



eCOMMONS

Loyola University Chicago
Loyola eCommons

DissertationsTheses and Dissertations

2010

Humanity in the Balance: The Relationship Between the Moral Law and the Promotion of the Moral World in Kant's Ethics

John J. Garcia
Loyola University Chicago

Recommended Citation

Garcia, John J., "Humanity in the Balance: The Relationship Between the Moral Law and the Promotion of the Moral World in Kant's Ethics" (2010). *Dissertations*. Paper 213.
http://ecommons.luc.edu/luc_diss/213

This Dissertation is brought to you for free and open access by the Theses and Dissertations at Loyola eCommons. It has been accepted for inclusion in Dissertations by an authorized administrator of Loyola eCommons. For more information, please contact ecommons@luc.edu.



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 License](https://creativecommons.org/licenses/by-nc-nd/3.0/).
Copyright © 2010 John J. Garcia

LOYOLA UNIVERSITY CHICAGO

HUMANITY IN THE BALANCE:
THE RELATIONSHIP OF THE MORAL LAW
TO THE PROMOTION OF THE MORAL WORLD
IN KANT'S ETHICS

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE GRADUATE SCHOOL
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

PROGRAM IN PHILOSOPHY

BY

JOHN J. GARCIA

CHICAGO, IL

MAY 2010

Copyright by John J. Garcia, 2010

All rights reserved.

ACKNOWLEDGEMENTS

I would like to thank all of the people who made this dissertation possible, beginning with the wonderful professors in the Philosophy department at Loyola University Chicago. Aside from the members of my dissertation committee I would also like to thank Dr. Adrian Peperzak, whose classes were an inspiration to me both as a scholar and as a teacher. I am of course deeply grateful to the members of my committee, Dr. Mark Waymack, Dr. Andrew Cutrofello, and Dr. Victoria Wike, for their time and for their thoughtful comments. I owe a profound debt of gratitude in particular to my committee chair and advisor Victoria Wike. Without her guidance and encouragement I would never have finished the project.

I would also like to thank Professor David Ozar for his friendship and for the opportunity to work for several years as a graduate assistant in the Loyola University of Chicago Center for Ethics. My time working at the Center for Ethics helped to make much of my graduate work financially possible. Additionally, it opened up several professional opportunities, for which I am deeply grateful.

Additional thanks to all of my colleagues at Harper College. I have been deeply touched by the support offered by so many. Again, with the support and encouragement of these friends and colleagues, I do not believe the project would have ever been completed.

Finally, thank you to my friends and family. Most of all, I would like to thank my wife Larissa. Without her I could do very little, let alone write a dissertation.

ABSTRACT

This dissertation deals with the tension between two seemingly divergent approaches to morality. On the one hand, there are those who take the view that morality concerns itself with the promotion of certain ends. This is a teleological or consequentialist view of ethics. On the other hand, we see thinkers who take the view that rationality or some other criteria provide us certain moral imperatives that may not be violated, regardless of our desire to bring about a particular end. Kant is usually depicted not only as a member of the latter camp, but indeed as the father of this approach. Occasionally these approaches to morality seem to be put into direct conflict with one another by cases in which one seems to face a choice between the promotion of ends and the adherence to certain moral rules.

One example of the supposed conflict between teleological concerns and formal requirements is famously depicted in the case of the murderer at the door. Many see Kant's approach to this case as one that causes us to act in a way that jars against our deep moral intuitions, and they take this to be a sign of a weakness in Kant's approach. As a result, thinkers such as Christine Korsgaard have attempted to read Kant in a way that sidesteps this conflict between teleology and form, arguing that the categorical imperative can be read in a way that allows us to lie to the murderer at the door. A view such as Korsgaard's is intriguing because it indicates a belief that we go wrong when we value a formal requirement such as the adherence to the dictates of rationality above the

desire to prevent a great injustice from occurring. This view is powerful, and it seems correct to me that, if our only reason to adhere to a conception of the moral law was to cling to a view of rationality, that this goal seems to pale in comparison to the desire to prevent great harms from occurring or to promote moral ends. Ultimately though, I think Korsgaard's approach fails.

I argue instead that Kant's ethical thought shows a deep concern for both teleological and formal considerations, and that a consideration of the relationship between these two aspects of his thought will help us make sense of his approach to cases such as the murderer at the door. It is the goal of my dissertation to present such an analysis. In my proposed dissertation I take the view that, far from interfering with the promotion of moral ends, Kant sees the formal requirements of morality as providing the only possible path to the highest end, a moral world. On my view, Kant's formal ethics and his teleology do not then represent stages in his thinking, or pieces of his thinking that stand at odds with one another; they are instead to be seen as two inseparable pieces of the same puzzle. A full understanding of each of these pieces of Kant's thought will show us that neither piece can make sense without the other.

Human beings have two sorts of ends: moral ends, which we set for ourselves, and natural ends, which aim at our happiness. Kant realized that obedience to the moral law was not important simply because it allowed us to be rationally consistent. He also saw such adherence to the moral law as the only sure path to the full realization of our humanity. It is important to remember that for Kant the realization of our humanity

involves a realization of the natural ends and rational ends for all human beings. So, Kant's project is much more 'cosmopolitan' than we often understand it to be.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	iii
ABSTRACT	v
CHAPTER ONE: INTRODUCTION	1
Korsgaard, the Categorical Imperative, and the Case of the Murderer at the Door	8
Korsgaard's Concern: The Confrontation with Evil	10
CHAPTER TWO: KORSGAARD'S CASE FOR THE NON-EQUIVALENCE OF THE FORMULATIONS OF THE CATEGORICAL IMPERATIVE	14
The Case for the Presumption of the Equivalency of the Formulas	14
Korsgaard's Casuistry in the Case of the Murderer at the Door	21
Korsgaard's 'Practical Contradiction' Interpretation	29
One Contradiction or Two?	33
Korsgaard on Rational Willing	36
CHAPTER THREE: KANT AND OBJECTIVE ENDS	44
The Ground vs. the Application of the Moral Law	50
Subjective Ends vs. the Objective End of Humanity as an End-In-Itself	56
Paton's View	60
Kant's Four Examples and the Connection between the Formula of Universal Law and the Formula of Humanity	70
Ends that Are at the Same Time Duties	83
CHAPTER FOUR: THE MORAL WORLD, MORAL PROGRESS, AND THE MORAL LAW	94
Kant's Conception of the Moral World and its Possibility in the World of Sense	101
Nature and History – Progress Toward the Highest Good	109
The Moral Will and the Moral Law	121
Problems with the Proportionality of Happiness and Virtue	128
Conclusion	137
CHAPTER FIVE: THE PLAUSIBILITY OF KANT'S APPROACH	138
Torture and the Contingency of Conditions vs. the Certainty of the Moral Law	141
Billy Budd and the Importance of Lawfulness	152
Health Care, Lying for Patients, and Reform	159
The Challenge from Feminist Ethics: Partiality vs. Impartiality	167

Korsgaard's Double Level Approach	173
CHAPTER SIX: CONCLUSION – THE PLACE OF HOPE IN AN IMPERFECT WORLD	190
BIBLIOGRAPHY	198
VITA	204

CHAPTER ONE

INTRODUCTION

This dissertation deals with the tension between two seemingly divergent approaches to morality. On the one hand, there are those who take the view that morality concerns itself with the promotion of certain ends. This is a teleological or consequentialist view of ethics. On the other hand, we see thinkers who take the view that rationality or some other criteria provide us certain moral imperatives that may not be violated, regardless of our desire to bring about a particular end. Kant is usually depicted not only as a member of the latter camp, but indeed as the father of this approach. Occasionally these approaches to morality seem to be put into direct conflict with one another by cases in which one seems to face a choice between the promotion of ends and the adherence to certain moral rules.

This issue has become increasingly important in our modern world, especially as a result of the relatively new ‘war on terror’. Many have begun to argue that, for example, torture may be justifiable in some cases, in the attempt to prevent harm from occurring to others. There are of course many consequentialist arguments both for and against the practice of torture. I find these kinds of arguments (regardless of the position they are used to defend) largely unsatisfying. The following example may help explain why. A couple of years ago I was at a conference and heard a panel discussion on the issue of torture. One panelist was defending the use of torture, or at least extremely

rigorous interrogation tactics, in at least some cases.¹ He told the following story as an example of his point.

A woman steps in to a store for a moment and leaves her car running with her infant inside. As soon as she enters the store, the car is stolen. The thief drives several miles and realizes that there is a young child in the car, panics, loots the car, and abandons it. The police apprehend the suspect not long after he abandoned the car, and are absolutely certain he is indeed the person who had taken the car. They interrogate him in an attempt to get him to admit to taking the car, but also, more importantly, to get him to reveal the whereabouts of the car so that they might save the child from dying due to the extreme summer heat. The suspect refuses to admit his guilt and therefore will not reveal the location of the car. After trying all other available methods (let's grant for the sake of the argument that this is true), one officer proceeds to beat the suspect until he gives up the location of the car and the child. The police get to the car and save the child, knowing that, had they arrived just a few minutes later, the child would likely have succumbed to the heat.

The panelist cited this example as the epitome of a case in which the actions of the officer could be justified. The child was saved. The abusive techniques had worked. It is important to remember that this case is largely immune from objections based on consequentialist concerns. They *knew* they had the right suspect, *knew* he was aware of the location of the car, and they did in fact save the child in jeopardy. However, I was

¹ This panel discussion took place at the 2006 annual meeting for the Association for Practical and Professional Ethics. I regret that I cannot remember the name of the panelist.

troubled by one question, which I asked him: “Would the actions of the officer still have been justified if they got to the car just *after* the child had succumbed to the heat?” To my surprise, the panelist responded that, in his view, the police officer’s actions would no longer have been justified. It simply seemed wrong to me then, and still does now, that the morality of a person’s actions could hinge on the amount of air in a car at any given time, which is something that is entirely out of the control of the person making the decision to engage in tactics we would normally find objectionable. The Kantian conclusion I kept coming back to was, either the act is wrong or it is not wrong; it cannot depend on contingent states of affairs.

Another example may help us to see that our desire to prevent harms from occurring, or to bring about certain good outcomes, might actually make a moral world more difficult to attain. Doctors often face the following dilemma when dealing with a patient who needs care but cannot afford it. The doctor might ask: should I lie on the insurance form in order to help ensure the patient gets the care he or she needs? The temptation is to say that we must ‘bend the rules’ to bring about a good outcome. However, one can see easily that, by engaging in such an act of deception, the doctor plays a small role in making a flawed insurance system worse. Rather than hope for and even work toward an improved system, the physician instead builds deception into the system. While this may help bring about some benefits for individuals, it actually makes improvement of the system more difficult². Kant, I will argue, understood this problem;

² I take up this example in more depth in Chapter Four of this dissertation.

this is one of the very reasons that we cannot let a consideration of consequences guide our view of morality. Yet, this answer might seem trite to many. In the previous example there is indeed a trade off. The physician forfeits the ability to help that particular patient and is left only with the hope that he or she can work within moral means to make the situation better in the future. I sympathize with those who would say that this is unrealistic view in the world in which we now live.

One could imagine situations when the stakes were indeed so high that it seems that anyone who still insists: “it’s just wrong, no matter what,” has lost touch with common sense. I actually agree with this judgment. If all we can say about the importance of adherence to the moral law is that the categorical imperative tells me it is wrong and therefore I may not do it, then we seem to abandon all concern for real decisions in favor of stubborn headed adherence to a formula for its own sake.

However, this dissertation is an attempt to show that Kant did not see morality this way. Adherence to the moral law is not only valuable for its own sake, but also because it is a way of seeing that rationality and rational agency are valuable for their own sake. Further, respect for the moral law and for rational agency is the only way to make progress toward a moral world possible. Respect for the moral law goes hand in hand with respect for persons and with the creation of a moral world. Only by seeing the extent of these connections can we see the real power of Kant’s approach to morality.

Kant himself took up the question of the alleged tension between teleological concerns and formal requirements in the case of the murderer at the door. It is interesting

to look at Kant's treatment of this situation in his essay, *On a Supposed Right to Lie from Philanthropy*, in part because the critic to whom he is responding, Benjamin Constant, bases his criticism of Kant's insistence that we may not lie to the murderer at the door on the claim that such adherence to a principle of truthfulness would make society untenable. (*Right to Lie* 425) This criticism reminds us of those who would say today that strict adherence to some set of moral rules causes us to lose touch with reality and allow unthinkable consequences. The nature of Kant's response to this criticism is insightful as well. Kant's response to Constant can be best captured in the passage in which he writes:

Truthfulness in statements that one cannot avoid is a human being's duty to everyone, however great the disadvantage to him or to another that might result from it; and although I do no wrong to him who unjustly compels me to make the statement if I falsify it, I nevertheless do wrong in the most essential part of duty *in general* by such falsification, which can therefore be called a lie ... that is, I bring it about, as far as I can, that statements (declarations) in general are not believed, and so too that all rights which are based on contracts come to nothing and lose their force; and this is a wrong inflicted upon humanity generally. (*Right to Lie* 426)

Note that Kant does not appeal to any notion of treating a person as an end. He even says that I "do no wrong to him who ...compels me to make the statement." Nor does Kant here refer to the issue of being able to will the universal practice of our maxim. This may seem startling to those who, even attempting to defend Kant's insistence that we may not lie, primarily point either to the fact that we fail to treat the person to whom we lie as an end or to the idea that lying entails some sort of contradictory or irrational willing. It may be tempting therefore to think that Kant is just off his game a bit in this essay.

(Indeed the whole example seems flawed by the fact that there is no such thing as a statement one cannot avoid, since I can always elect to remain silent and face any consequences that would result from this silence.) However, I would suggest that Kant gives us an often overlooked insight in the passage noted above.

Kant notes that our duty of truthfulness is instead based on the fact that, by lying, I play a part in willing a world in which we can not know with certainty what is truth and what is falsity, and therefore I undermine “rights which are based on contracts” and therefore harm “humanity generally”. The goal of this dissertation will be to show that this statement actually points to a central thought in Kant’s ethical work. Kant understood the connection between adherence to the moral law and the possibility of a moral world.

My attempt to understand Kant in a way that shows that he understood the connection between the moral law and progress toward a moral world differs in many important regards from other authors who have tried to find a balance between teleology and deontology in Kant. I will be talking at length about one author’s approach in particular, namely the approach taken by Christine Korsgaard in her book *Creating the Kingdom of Ends*. While reading Korsgaard’s text, I felt myself drawn to her project because she attempts there to strike a balance between adhering to the commands of the categorical imperative and making choices that help promote a Kingdom of Ends. However, her approach is very different from the one I will take, and I will offer arguments that her approach will not ultimately work.

Korsgaard takes the approach of arguing that the various formulations of the categorical imperative in fact lead to different results, and that the Formula of Universal Law will allow us to lie to the murderer at the door while the Formula of Humanity and the Formula of a Kingdom of Ends will not. This is critical for Korsgaard's argument because she believes, as I do, that we ought to see the creation of a moral world, or a Kingdom of Ends, as the ultimate goal of moral action. She thus argues that we will sometimes have a choice between the less stringent Formula of Universal Law and the other more demanding formulations. When we have such a choice, Korsgaard argues that we ought to prefer the more lenient formulation, since adherence to the more stringent Formula of Humanity may actually interfere with our ability to deal with evil and to promote certain ends.

A view such as Korsgaard's is intriguing because it tries to remain true to a Kantian approach to morality while it at the same time indicates a belief that we go wrong when we value a formal requirement such as the adherence to the dictates of rationality above the desire to prevent a great injustice from occurring. This view is powerful and it seems that, if our only reason to adhere to a conception of the moral law was to cling to a view of rationality, Korsgaard would be right to suggest that this goal seems to pale in comparison to the desire to prevent great harms from occurring or to promote moral ends. Korsgaard tries to make room within Kant's thought to loosen the requirements put on us by the categorical imperative, and she further argues that this is

critical if we are to preserve our goal of creating a Kingdom of Ends when it seems as if the very requirements of the categorical imperative threaten to make this more difficult.

I will argue that Korsgaard is incorrect when she argues that the formulations of the categorical imperative give different results in practice and that we are thus free to choose which formulation to follow when they conflict. I will further argue that, if our choice really is one between following the strict commands of the categorical imperative and promoting certain ends, Kant was right to insist that the dictates of duty must take precedence if moral action is to retain any meaning. I will further attempt to show that Kant indeed saw the adherence to the commands of the categorical imperative as a necessary condition for the attainment of a Kingdom of Ends. However, in order to appreciate the difference between my approach and the one taken by Korsgaard, it may be useful to look briefly at her approach to the case of the murderer at the door.

Korsgaard, the Categorical Imperative, and the Case of the Murderer at the Door

Korsgaard treats the case of the murderer at the door as part of her book *Creating the Kingdom of Ends*, and it plays an important role in her argument that the formulations of the categorical imperative are indeed distinct and that we ought to follow the less stringent Formula of Universal Law rather than the more restrictive Formula of Humanity.

Kant insists that a lie is always wrong, even in a case in which we are asked to choose between being truthful to a murderer at our door and lying to the murderer in an attempt to save a life. Christine Korsgaard points out the paradox caused by such a case

when she notes that Kant's insistence that a lie is always wrong implies that, "morality itself sometimes allows or even requires us to do something that from an ideal perspective is wrong" (Korsgaard 135). Simply put, it seems as if Kant's approach to morality forces us to act immorally.

Readers of Kant have reacted to his treatment of the case of the murderer at the door in various ways. Korsgaard does a nice job articulating the approaches taken both by those who are sympathetic to Kant and those that are unsympathetic to him. Of the unsympathetic readers, Korsgaard writes,

Unsympathetic readers are inclined to take [Kant's claims that we cannot lie to the murderer at the door] as evidence of the horrifying conclusions to which Kant was led by his notion that the necessity in duty is a rational necessity – as if Kant was clinging to a logical point in the teeth of moral decency. Such readers take these conclusions as a defeat for Kant's ethics.... (Korsgaard 134)

Korsgaard then contrasts this view with the one taken by readers who are sympathetic to Kant's overall approach to morality. She notes:

Sympathetic readers are likely to argue that Kant mistook the implications of his own theory, and to try to show that by careful construction and accurate testing of the maxim on which this liar acts, Kant's conclusions can be blocked by his own procedures. (Korsgaard 134)

It seems then, that the case of the murderer at the door tempts readers to either take the view that Kant is committed to his claim that such lies are impermissible, and thus to use this as a reason to dismiss his whole approach to ethics as impractical, or to take the view that the categorical imperative is properly interpreted as allowing such lies, and that Kant misunderstood the implications of his own claims. The common feature of both these approaches is that they both treat the categorical imperative as *one* moral principle with

one set of results. However, Korsgaard attempts to take a third approach to cases such as the murderer at the door. She does this by arguing that:

when the case is treated under the Formula of Universal Law, this particular lie can be shown to be permissible[However]when the case is treated from the perspective of the Formulas of Humanity and the Kingdom of Ends, it becomes clear why Kant *is* committed to the view that lying is wrong in every case (Korsgaard 134-5).

For Korsgaard then, when it comes to the case of the murderer at the door, “the Formula of Universal Law and the Formula of Humanity give us different results” (Korsgaard 144). Korsgaard takes the view that we go wrong when we say that there is only one answer to the question: Does the categorical imperative allow us to lie to a murderer at the door? In order to understand the importance of this claim, that the formulations of the categorical imperative yield different results, we must first understand Korsgaard’s motivation for making the claim.

Korsgaard’s Concern: The Confrontation with Evil

In the conclusion of Chapter Five of *Creating the Kingdom of Ends* Korsgaard makes it clear why it is that she argues that the formulas of the categorical imperative yield different results in practice. She urges us to see the Formula of Humanity as providing a standard that best serves as an ideal toward which to strive but which can be unrealistically demanding in some cases. Yet, we need not give up the moral guidance of the categorical imperative entirely, since we could still refer to the less strenuous Formula of Universal Law. She writes:

The Formula of Humanity and its corollary, the vision of a Kingdom of Ends, provide an ideal to live up to in daily life as well as a long-term political and

moral goal for humanity. But it is not feasible always to live up to this ideal, and where attempts to live up to it would make you a tool for evil, you should not do so. In evil circumstances, and only then, the Kingdom of Ends can become a goal to seek rather than an ideal to live up to. ... And even in the worst circumstances there is always the Formula of Universal Law, telling us what we must not in any case do. ... The Formula of Universal Law provides the point at which morality becomes uncompromising. (Korsgaard 153 – 154)

Korsgaard thus intends to use the Formula of Universal Law as the limit of our actions, and so maintains an approach to morality that she can comfortably say is truly Kantian. However, she argues that we should consider the Formula of Humanity as an ideal that we may depart from when dealing with evil. She claims that, “[i]n such cases, we can say that the Formula of Humanity is inapplicable because it is not designed for use when dealing with evil” (Korsgaard 151). This approach amounts to what Korsgaard calls a “Kantian double-level theory” (Korsgaard 151). She bases this double-level theory off of a Rawlsian distinction between ideal and non-Ideal theories.

Korsgaard explains that in Rawls’ *A Theory of Justice*, he had proposed a distinction between Ideal and non-Ideal theories (Korsgaard 147). Michael Phillips describes Rawls’ distinction in the following way:

Very roughly, Ideal Theory attempts to describe those principles for the design of institutions and the conduct of persons that would be appropriate to a morally and politically ideal order, while non-Ideal Theory concerns itself with the principles that would be appropriate for these purposes under less perfect conditions. (Phillips, 551)

Korsgaard uses this distinction to explain the difference between Rawls’ general conception of justice and his special conception of justice.

According to Korsgaard, Rawls' special conception of justice involves the equal distribution of liberty and opportunity, while the general conception of justice allows an unequal distribution if it is to the advantage of everyone, especially those most disadvantaged (Korsgaard 147). She continues on to say that, given specific non-Ideal conditions, "the special conception becomes a goal, rather than an ideal to live up to, we are to work toward conditions in which it is feasible" (Korsgaard 148). Note however that even when we leave the special conception of justice, we still have some guidance from the less strict general conception. Korsgaard puts it this way: "Rawls' special conception of justice is a stricter version of the egalitarian idea embodied in his general conception" (Korsgaard 151). This sort of double-level approach, in which we can abandon a more strict principle yet still remain bound by a less strict principle that still expresses the same idea, is what Korsgaard attempts to apply to Kant's approach to morality when she argues that we can treat the Formula of Humanity as a goal while still making sure to keep within the bounds of the Formula of Universal Law.

Korsgaard notes that for Kant, "we are always to act as if we are living in a Kingdom of Ends, regardless of possible disastrous results" (Korsgaard 149). The problem, of course, is that we do not live in such a world. Human beings often act in opposition to the creation of the Kingdom of Ends, and nature sometimes seems to do so as well. Korsgaard finds this standard unrealistic. She even goes further and calls it "grotesque simply to say that I have done my part by telling the truth and the bad results

are not my responsibility” (Korsgaard 150). Korsgaard is not here accusing Kant of cold-heartedness. She notes that:

Kant is by no means dismissive toward the distressing problems caused by the evil conduct of other human beings and the unfriendliness of nature to human ideals. ... He finds in [these problems] grounds for a morally motivated belief in God. Our rational motive for belief in a moral author of the world derives from a rational need for grounds for hope that these problems will be resolved. (Korsgaard 149)

Korsgaard is correct to say that Kant believes that a rational belief in God is a necessary component for our faith in the eventual attainment of the Highest Good. However, she is incorrect when she says that this is Kant’s sole solution to the problems of evil doers and injustices in the world. Kant’s view of morality commits him to the view instead that we must hold fast to the belief that our commitment to morality is a large part of making the Kingdom of Ends possible. His entire approach is an argument against letting empirical considerations or desired outcomes serve as the criteria by which we determine what is right.

I wish not only to suggest that Kant would reject Korsgaard’s adjustment to his thought, but that he is right to do so. Korsgaard presents us with an interesting dilemma: can Kant present a satisfactory way to deal with evil? However, her solution is untenable. I wish to show that an emphasis instead on the unity (but not identity) of the formulations is a better solution to the alleged tension between our desire to promote a more ideal world and a real commitment to the demands of the categorical imperative. The first step in this process is to examine the claim that the formulas of the categorical imperative are indeed non-equivalent.

CHAPTER TWO

KORSGAARD'S CASE FOR THE NON-EQUIVALENCE OF THE FORMULATIONS OF THE CATEGORICAL IMPERATIVE

Korsgaard's argument that the formulas of the categorical imperative are non-equivalent raises several important issues. In this chapter I will begin by looking at some preliminary reasons why it may be wise to assume that the burden of proof rests on those who would argue that the formulas are nonequivalent. Then I will examine Korsgaard's casuistry in the case of the murderer at the door in order to see if it works as an argument that the formulas indeed yield different results. Finally, I will examine Korsgaard's notion of what sort of contradiction is involved in willing unacceptable maxims in Kant's Formula of Universal Law. This last issue affects one's casuistry in cases such as the murderer at the door, but it is also important for another reason. It is a critical issue because it introduces the question of just what is involved in rational willing. Once we come to an understanding of what it is that a rational agent can and cannot will, we will begin to see the nature of the relationship between the sort of formalism that seems to appear in the Formula of Universal Law and the teleological considerations that appear elsewhere in Kant's thought.

The Case for the Presumption of Equivalence between the Formulas

As I have already said, Korsgaard argues that the formulations of the categorical imperative are nonequivalent because they yield different results, particularly in the case

of the murderer at the door. John E. Atwell however expresses the concern that critics often argue for the nonequivalence of the formulations of the categorical imperative by offering their own interpretation of one of the formulas. Atwell then suggests that such critics “may have shown only that Kant misapplied or misunderstood one of the principles, rather than that the principles themselves are nonequivalent” (Atwell 275). Atwell continues on to give a few examples, noting for instance that Marcus Singer, in his work *Generalization in Ethics*, “gives his own interpretation of the first principle and shows ... that it does not imply that lying is always wrong. But then, noting that Kant regarded every case of lying as treating someone merely as a means ... concludes that the two principles are not equivalent” (Atwell 276). We must then make sure that Korsgaard is not engaging in this same sort of discussion which would amount to the claim that Kant *ought* to have seen the principles as nonequivalent even if he did see them as equivalent.

On the other hand, Onora O’Neill cautions against the temptation too arrive too quickly at the belief that the formulations are equivalent. She notes that there is a distinction between ‘extensional equivalence’ and ‘intensional equivalence’ and that true equivalence requires that both of these standards be met (O’Neill 364). She explains the distinction in the following way:

The two formulations might be equivalent in that both classify maxims, and derivatively the acts that conform to or violate those maxims in the same ways: they might be simply extensionally equivalent.... Alternatively ... [the formulations] might be intensionally equivalent, if it could be shown not merely that they in fact yield the same results, but that this follows from the nature of the formulations. (O’Neill 346)

O'Neill is here making the point that the overlap in results between two formulations cannot be a matter of serendipity. The formulas must in fact *logically* entail identical outcomes when applied as criteria for action. This is because, "if we want insight into why these formulations are both versions of the supreme principle of morality we will need [it] to be shown not merely *that* they yield the same results, but *why* they do so" (O'Neill 346). This comment is extremely important, as it suggests that there is good reason to be hopeful that the formulations are indeed equivalent. O'Neill's comment points to the fact that the formulations of the categorical imperative are not a series of guides to action, but are allegedly expressions of the same principle of morality.

O'Neill elaborates on the issue of the importance of belief that the formulas are equivalent. She notes that nonequivalence between the Formula of Universal Law and the Formula of Humanity¹ would jeopardize the legitimacy of the appeals to the Formula of Humanity by those who see this formula as more fruitful. She makes this point when she writes:

Most of the arguments or argument sketches that he [Kant] provides for the supreme principle of morality lead us to (at least towards) ...[the Formula of Universal Law]; yet much that he and many of his admirers (and even of his critics) find attractive and significant in guiding moral reflection derives from the ...[Formula of Humanity]. It is the ideal of treating people as ends and avoiding using them as means, not the ideal of acting on universalizable principles, that has become part of our culture. If the attractive idea of treating people as ends and never as means may not be groundable by Kantian arguments, while the charges of rigourism and formalism which are perennially leveled against ...[the Formula of Universal Law] may lead us to conclude that even if Kantian arguments show that this is the supreme principle of morality, still we have not discovered a principle that can help us lead our lives. (O'Neill 342)

¹ O'Neill uses "Formula of Ends in Themselves" but I am using Korsgaard's language here, as my argument in this chapter deals primarily with Korsgaard's treatment of the formulations.

It might seem as if O'Neill's worry that we place the Formula of Humanity in jeopardy if we accept the conclusion that the formulas are nonequivalent would not be a concern to Korsgaard, since it is after all the Formula of Humanity that she advises us to give up when it overly confines our actions in the face of evil. However, O'Neill's remark is important for the reason that it shows that any attempt to argue for the nonequivalence of the formulations comes at a price. This affects Korsgaard's view because, while it may be helpful to see the formulas as separate so that we have some latitude in our actions when confronting evil, we would have to at the same time surrender all the benefits that come from seeing the formulas as equivalent. O'Neill's comments point to the fact that Kant's argument in the *Groundwork* only makes sense "if there is some reading of the formulations under which the claimed equivalences hold" (O'Neill 353). This serves by itself as a *prima facie* reason to put the burden of proof on those who would interpret the formulas as nonequivalent.

We can also add another reason to take the view that the default position should be one of seeing the formulas as equivalent. This reason comes in the form of Kant's own language. Kant notes that the various formulations represent "only so many formulae of the very same law, and any one of them unites the other two in it" (*Gr* 436). This quote is admittedly frustrating, since Kant then goes on to say: "There is nevertheless, a difference among them" (*Gr* 436).² If we take Kant's words together with

² O'Neill makes this same point on page 342 of "Universal Laws and Ends-in-Themselves" and attempts to explain this tension in the remainder of her article. I will take up this point again later on, and will make use of some of her points in doing so. Suffice to say, I agree with O'Neill that it is possible to have

the above considerations about the importance of an equivalence between the formulations to Kant's project we are able to see that any reading sustaining the equivalence of the formulations "gains some support from the fact that Kant claims they are equivalent and that it is vital to his argument that they be so" (O'Neill 343).

So, as we turn to look at Korsgaard's argument that the formulations are not equivalent, we must keep the forgoing considerations in mind. We must ask, with Atwell, whether Korsgaard is giving Kant's view of each formulation or whether she is offering a new interpretation of one or more of the formulations. If she is doing the latter, then it seems that this would be insufficient to truly show nonequivalence. Additionally, we must ask, with O'Neill, whether the arguments for the nonequivalence are sufficient to trump considerations in favor of the view that the formulations are equivalent.

Finally, it is important to point out one more consideration regarding the importance of the move of leaving the Formula of Universal Law intact as a guide for our actions. Let us assume for now that Korsgaard is correct that the Formula of Humanity is too strict and is indeed unfit for dealing with evil. If the formulas of the categorical imperative were equivalent, or, even if non-equivalent, if Korsgaard had advised us to abandon the categorical imperative entirely as a guide for actions, we would be left without real guidance when it comes to the issue of how we are to deal with evil. It is dangerous and incorrect to view Korsgaard as saying that, when dealing with evil, we are

equivalence in terms of the implications for ethical evaluation without having a complete identity between the formulations. Indeed my argument will try to show the importance of the view that the formulations can be complementary even if they are functionally equivalent.

beyond the scope of the categorical imperative and must do whatever we believe will be most effective for ultimately defeating evil and promoting a kingdom of ends. It does not take much to realize that such a position would be hardly worth considering, at least not for anyone who calls themselves a Kantian. Currently, in the United States, much deception, coercion and cruelty have been excused in the name of dealing with the evils of terrorism. If Korsgaard had argued that we ought to sometimes abandon the categorical imperative altogether because it is too strict, this would amount to a view that we ought to follow a Kantian conception of morality, but only to a point, and that we should abandon it when the going gets tough³.

However, by separating the formulations, Korsgaard takes a step that weakens the force of the categorical imperative and begins to allow our ends to *determine* our duty. Heiner Bielefeldt explains why the view that the formulations of the categorical imperative are non-equivalent in fact undermines the concept of a categorical imperative. He writes, “In such a conceptualization, the very uniqueness – and thus – also the strictness – of the categorical imperative would actually be lost” (Bielefeldt 56). So, if we were to follow Korsgaard, the categorical imperative would be replaced by a set of different (or even antagonistic) moral principles, none of which would any longer have the unconditioned status of the categorical imperative. Positively speaking, the

³ One must wonder however what Korsgaard would have us do if in fact the formulas were equivalent. If so, and if her avenue was closed to her, should we prioritize the goal of a kingdom of ends and abandon Kant’s very tool for making that kingdom a reality, or should we stay true to Kant’s vision of the categorical imperative and risk the feeling that we have let evil get the better of us due to our allegiance to rationalist principles? I take this issue up in the final chapter.

categorical imperative must be a single imperative in order to have its categorical authority (Bielefeldt 56).

Even though Korsgaard maintains that the formulations still express the same idea, even though they yield different results, the fact is that they do in fact yield different results in her account and thus compete with one another. Bielefeldt's comments lead to an important point. When formulations compete, how are we to decide which formula to follow? Korsgaard has given her reasons to choose the Formula of Universal Law, but it is clear that she is here no longer interpreting Kant but suggesting a revision to his thinking. However, we must ask whether this revision is for the best.

Korsgaard puts the creation of a Kingdom of Ends as the goal of moral action and argues that we ought to view the Formula of Humanity and the Formula of the Kingdom of Ends as too unrealistic (in that they are too demanding) when they seem to make the very realization of the Kingdom of Ends less possible. I believe that this amounts to a sort of lack of faith in Kant's view of morality.

I have attempted to argue that there are several reasons we ought not to follow Korsgaard's claim that the formulas of the categorical imperative are nonequivalent. However, all of my arguments so far have pointed solely to the repercussions of such a view and not to a consideration of the actual use of the formulas as guides to action. With that in mind, let's now look at her treatment of this case.

Korsgaard's Casuistry in the Case of the Murderer at the Door

Korsgaard argues that the Formula of Universal Law would allow us to lie to the murderer at the door while the Formula of Humanity would not allow such a lie. I consider her claim that the Formula of Humanity would not permit the lie to be non-controversial and will focus my attention here on her claim that the Formula of Universal Law would permit lying in the case of the murderer at the door. Korsgaard's claim that the Formula of Universal Law would permit such a lie rests on her casuistry of the case, and her casuistry rests on a further assumption regarding the nature of a contradiction involved in willing untenable maxims. I will treat her casuistry here and then go on to deal in depth with her discussion of the contradictions involved in willing unacceptable maxims. However, her discussion of the nature of such contradictions needs to be briefly mentioned here, as it fuels her casuistry.

Korsgaard begins her analysis by noting that Kant's formula of universal law puts us in the position of having to figure out "what you can will *without contradiction*" (Korsgaard 77). Korsgaard then notices that there are various views about the ways in which a maxim might be said to come into contradiction with itself (Korsgaard 80), and she adopts what she refers to as a 'practical contradiction interpretation' and goes on to assess the case of the murderer at the door under this interpretation.⁴ Although I believe that there are problems with Korsgaard's view, that the sort of contradiction involved in a maxim that cannot be universalized is what she calls a practical contradiction, I will

⁴ Korsgaard's argument for the practical contradiction interpretation is in Chapter Three of *Creating the Kingdom of Ends*. She then treats the case of the murderer at the door, using this interpretation, in Chapter Five of the same work.

postpone that question for now and instead will begin simply by examining the casuistry of Korsgaard's analysis and application of the practical contradiction interpretation.

Korsgaard summarizes her view, that the contradiction we are to avoid when we will a maxim as a universal law of nature is a practical one, when she writes: "The universalized maxim contradicts itself when the efficacy of the action as a means for achieving its purpose would be undermined by its universal practice" (Korsgaard 135). Korsgaard then applies this to the case of lying, saying that "lies are usually efficacious in achieving their purposes because they deceive, but if they were universally practiced they would not deceive" (Korsgaard 136). It seems that, so far, Korsgaard does not deviate from the conclusion that any maxim that would involve willing a lie would seem to be forbidden by the formula of universal law. However, she then presents the case of the murderer at the door in such a way that lying seems to become permissible.

Korsgaard begins her depiction by stating that, "there is probably already deception in the case" (Korsgaard 136). Korsgaard claims that this deception exists because a potential murderer would not be honest about his intentions, "in any event, this is not how I [Korsgaard] shall imagine the case" (Korsgaard 136). It is important to understand that Korsgaard is not arguing that it is the immorality of the murderer's deception that makes our lie permissible. She is not making the same argument to which Kant is responding in, *On A Supposed Right to Lie From Philanthropy*. In that work he criticizes those who hold the view that our duty to tell the truth only applies to those who

have a right to it.⁵ It is obvious of course, that Korsgaard must be taking a different approach. For, if her claim were that we could be exempted from our moral obligations by the evil of others, then his deception would be irrelevant, for he already had murderous intentions. Although Korsgaard believes that the murderer has, “placed himself in a morally unprotected position by his own deception,” (Korsgaard 137) this is not because morality no longer binds us at all.⁶ Rather, Korsgaard’s claim is that the murderer’s deception makes it possible for us to universalize a maxim of lying without entangling ourselves in a contradiction. However, this claim depends on another nuance Korsgaard adds to her depiction of the situation.

Korsgaard has already stipulated that she will treat the case of the murderer at the door as one in which the murderer is himself engaging in deception. She then goes on to specify that, “A murderer who expects to conduct his business by asking questions must suppose that you do not know who he is or what he has in mind” (Korsgaard 136). Given the qualifications Korsgaard has made to the case, the lie then becomes permissible because:

The lie will be efficacious even if universally practiced. But the reason it will be efficacious is rather odd: it is because the murderer supposes you do not know the circumstances you are in – that is, that you do not know that you are addressing a murderer – and so does not conclude from the fact that people in those circumstances always lie that you will lie. (Korsgaard 136)

⁵ Kant’s response is to an essay by Benjamin Constant. Kant summarizes Constant’s claim in section 426 of the *Supposed Right to Lie From Philanthropy*.

⁶ This should be clear from the fact that Korsgaard holds the view that the formula of the Kingdom of Ends and the Formula of Humanity would not permit our lie, despite the murderer’s deception.

The situation then, as Korsgaard depicts it, is as follows: First, the murderer at the door engages in an act of deception by keeping his intentions secret. Second, you are somehow aware of his intentions anyway. Third, the murderer does not know that you are on to him. For Korsgaard then, I am able to will my maxim as a universal law (assuming Korsgaard's interpretation of the reasons why a maxim cannot be universalized) because even if everyone lied in the circumstance she describes, the lies would never cease to be efficacious.

The lie would be efficacious because of the subtleties Korsgaard builds into the case. Because the murderer is unaware that I know of his murderous intentions, he will not expect that I will lie. So, even if this situation were to be repeated time and again, the murderer would never have any reason to assume that he has been told a lie. Lies are efficacious until they cause others to doubt our veracity, and on Korsgaard's account of the case, the murderer would never have any reason to doubt our responses to his questions.

Marcus Singer also takes this approach in his essay "The Categorical Imperative" which deals with one of the issues with the question of whether Kant's categorical imperative really implies the sort of absolutism that is normally attributed to him. Singer comments: "the view that certain moral rules hold absolutely or in all circumstances, simply does not follow from Kant's first moral principle" (Singer 583). He then uses the example of the murderer at the door to support this claim.

Singer, like Korsgaard, supposes that the reason a maxim of lying (or any maxim) would allegedly be rejected by Kant's formula of universal law is because the universal practice of the lie would be self-defeating. He then argues, as does Korsgaard, that the case of the murderer at the door presents a circumstance wherein the maxim would not be self-defeating if made universal. He makes this point when he says: "the lie would be self-defeating – the murderer would fail to believe it – only if he knew what the circumstances were, that is to say only if he knew that his victim was in the house" (Singer 587). Korsgaard and Singer thus seem to have found an interesting 'loophole' in Kant's prohibition against lying. However, if we examine their reasoning more closely, we will see that it has some troubling implications.

If the wrongness of lying rests on the fact that the universal practice of lying would undermine itself, Kant's entire moral project begins to look suspect. Ignoring the criticism that such a view seems to open Kant up to the charges that he is in fact a closet rule-utilitarian, there are other concerns that are prompted by Korsgaard's interpretation. On her explication, Kant's example of the wrongness of making a lying promise rests not on the fact that a lie has been told but that a false promise has been made. (Of course all false promises are also lies, but it's important to remember that not all lies are false promises.) If we realize that the lie specifically involves a *promise* we can see important ramifications. A false promise becomes wrong because it will inevitably become uncovered. I can't make false promises without the person to whom I have made the promise discovering my lie. So, clearly, such false promises cannot be universally

practiced because such a practice would indeed undermine itself. However, the situation is different with any lie that will not be uncovered. Korsgaard acknowledges one unusual outcome of this interpretation. She notes that her analysis allows us to lie not just to potential murderers but also to “liars whose purposes are good” (Korsgaard 145).

Korsgaard presents the following case, which I quote below at length:

Here is a case: suppose someone comes to your door and pretends to be taking a survey of some sort. In fact, the person is a philanthropist who wants to give his money to people who meet certain criteria, and this is his way of discovering appropriate objects for his beneficence. As it happens, you know what is up. By lying you could get some money, although you do not in fact meet his criteria.... By my argument it is permissible to lie in this case. The philanthropist, like the murderer, has put himself in a morally unprotected position by his own deception. (Korsgaard 145)

Korsgaard’s analysis thus ironically suggests a supposed right to lie in order to *receive* philanthropy. Beyond this, as I have said already, *any* lie that would remain uncovered would be permissible under the interpretation offered by Singer and Korsgaard. What is oddest about Korsgaard’s conclusion, that her analysis of the Formula of Universal Law commits us to the view that we may lie even when our own intentions are less than honorable, is that she does not see any problem with this view. It is important to remember that Korsgaard is not claiming to be offering an adjustment to Kant’s thought. She is claiming instead that Kant’s own thinking necessitates the conclusions she reaches. But, to put the matter simply, it seems preposterous to suggest that Kant would endorse the view that we may lie for our own gain. In fact I can envision no counter-example that would more clearly show that Korsgaard’s analysis must have gone wrong somewhere. In the end, Korsgaard twists Kant’s own thinking in order to try to save him from a

conclusion she believes is unacceptable: that we cannot lie to prevent evil from occurring. However, in her attempt to do so, she has led us to an even more absurd conclusion: that we may lie for our own benefit.

It might be said here that the objections I have raised against the view offered by Korsgaard and Singer hold regardless of one's view of the content of maxims. Singer argues that maxims are not general rules intended to guide actions independently of context, and thus stands in some opposition to authors who have argued that maxims are indeed far more general. He writes that the categorical imperative "must always be applied to an action considered as taking place in certain circumstances, or for a certain purpose. The proof of this is that a reference to the circumstances and purpose of an action is necessarily involved in the "maxim" of the action" (Singer 590). He goes on to say that "Kant overlooked the fact that lying merely for one's own personal convenience and lying in order to save the life of some innocent person are two different sorts of actions, actions whose maxims are quite different" (Singer 590). Singer's point is appealing, but it does not amount to an argument that the categorical imperative would allow us to lie if our intentions were altruistic. Indeed his only argument offered that we may lie to the murderer at the door under the Formula of Universal Law hinges not on our ends but on the fact that the lie would be believed. As I have already shown, if this is the reason the lie is permissible, then it leads us to several unacceptable and counter-intuitive conclusions.

However, Singer does raise an important issue. Kant does include a purpose in some of the maxims he uses in the *Groundwork*. He indeed does say that the maxim in the false promising example is the following: “when I believe myself to be in need of money I shall borrow money and promise to repay it, even though I know that this will never happen” (*Gr* 422). It does not seem that Kant needs to stipulate the reason for the false promise is that the person considering making the maxim would be in need of money, for the very act of making false promises seems contradictory, without mention of the reason why the promise was made. I will argue later on that Kant’s mention of purposes or ends is indeed important for understanding his ethical thought as a whole. However, as we have seen above, Singer’s claim, that the mention of the ends in view is important to the casuistry in the case of the murderer at the door, cannot be supported.

I have argued that there ought to be a presumption toward the view that the formulations of the categorical imperative are equivalent, and I have further argued that Korsgaard’s casuistry leads to counter-intuitive results. However, I believe that her approach cannot ultimately work primarily because her casuistry is based on a view of rational willing that is incorrect. Our view of what is to count as rational willing impacts our understanding of why it is that maxims cannot be successfully willed as universal laws, and it is our view of what causes maxims to fail that will always guide our casuistry when treating cases under the Formula of Universal Law. So, it is important to now look at Korsgaard’s understanding of the sort of contradiction involved in willing maxims that cannot be universalized.

Korsgaard's 'Practical Contradiction' Interpretation

Any attempt to use the categorical imperative's Formula of Universal Law as a principle for evaluating the morality of actions depends on a particular conception of what it is that renders a maxim contradictory when one attempts to will it as a universal law. Kant tells us that there are indeed two tests maxims may be subjected to when we ask whether they can be willed as universal laws. Kant explains the difference between the two when he writes:

Some actions are so constituted that their maxims cannot even be *thought* without contradiction as a universal law of nature, far less could one *will* as what *should* become such. In the case of others this internal impossibility is indeed not to be found, but it is still impossible to *will* that their maxim should be raised to the universality of a law of nature because such a will would contradict itself. (*Gr* 424)

These two tests mentioned by Kant have since come to be known as the contradiction in conception test and the contradiction in the will test, respectively.

Kant associates contradictions in conception with "strict or narrower (unremitting) duty" (*Gr* 424) and contradictions in the will with "wide (meritorious) duty" (*Gr* 424). Kant gives four examples of the Formula of Universal Law in action: a person contemplating suicide, a person considering making a false promise, a person contemplating neglecting his or her own talents, and a person contemplating the withholding of aid to others. These examples are intended to respectively represent perfect duties to ourselves, perfect duties to others, wide or meritorious duties to ourselves, and wide or meritorious duties to others. This delineation thus puts a case such as lying (even to a would-be murderer at the door) in the category of perfect duties

to others, and as such, as a duty to be treated under the contradiction in conception test. It would thus seem as if we have a simple criterion to decide whether or not the Formula of Universal Law would permit us to lie to the murderer at the door. If we can will such a maxim as a universal law without causing a contradiction in conception, then it seems like the maxim would be permissible. However, this all depends on what we mean when we say “contradiction in conception”.

Korsgaard offers her view of just what is involved in such contradictions when she writes:

On this interpretation, the contradiction is that your maxim would be self-defeating if universalized: your action would become ineffectual for the achievement of your purpose if everyone (tried to) use it for that purpose. Since you propose to use that action for that purpose at the same time as you propose to universalize the maxim, you in effect will the thwarting of your own purpose. (Korsgaard 78)

The important feature of Korsgaard’s interpretation is that it focuses on the impact the universalization of one’s maxim has on the efficacy of one’s own purpose. This sits in contrast to authors who focus instead on the logical impossibility of willing one’s maxim to be a universal law.

Korsgaard counts authors such as Dietrichson, Kemp and Wood ⁷ among those who see “something like a logical impossibility in the universalization of the maxim, or in the system of nature in which the maxim is a natural law: if the maxim were universalized, the action or policy it proposes would be inconceivable” (Korsgaard 78). Korsgaard refers to this view as a “Logical Contradiction Interpretation” (Korsgaard 78).

⁷ See page 81, and the footnote for Korsgaard’s exact references to these authors.

She then offers a summary of how such an approach would handle the case of false promising when she writes, “we are imagining a world in which the agent and everyone with his purpose is making a certain sort of promise, but also a world in which there is no such thing” (Korsgaard 82). Indeed Kant’s language in his example of a maxim of false promising makes it very difficult to determine whether he sees a practical contradiction or a logical contradiction at work. The language that causes the confusion occurs when Kant writes:

For, the universality of a law that everyone, when he believes himself to be in need, could promise whatever he pleases with the intention of not keeping it would make the promise and the end one might have in it impossible, since no one would believe what was promised him but would laugh at all such expressions as vain pretenses. (*Gr* 422)

When Kant refers to the promise itself becoming impossible, this seems to suggest a logical contradiction in which the practice of promising becomes nonsense. However, when, in the same sentence, he refers to *the end contained in the promise* becoming impossible, we see a possible reason to think in terms of a practical contradiction in which the end of the person making the false promise is thwarted by the fact that such false promises have been universalized. It thus seems that this issue cannot be settled by reference to the textual evidence in Kant’s own writing.

The difference between these two views is, however, critical. This is because Korsgaard uses her Practical Contradiction Interpretation (Korsgaard 78) as the foundation for her casuistry in the case of the murderer at the door, and thus it is only using that particular interpretation that she was able to argue that the Formula of

Universal Law would permit such a lie and that it is thus non-equivalent to the Formula of Humanity (since the lie would clearly be prohibited under that formulation). In contrast, I will show that a view that looked for a logical contradiction in the contradiction in conception test would find the maxim of lying to a murderer at the door unacceptable. Further, I will suggest that Korsgaard's interpretation that the type of contradiction involved in willing an unacceptable maxim is a practical one, is an untenable view.

Korsgaard's view implies a specific notion of rational willing, one which I believe is incorrect. In this section, I will try to offer reasons to reject her view, and in the next chapter I will offer my own interpretations as to Kant's view of rational willing. Our understanding of what is and what is not rational to will is crucial for my larger project. This is because such an understanding will help us see the connection between the Formula of Universal Law and the other formulations of the categorical imperative and to see the tie between the moral law and teleology in Kant's ethical thought. It is clear however, that we will need to appeal to criteria beyond Kant's own work to determine which interpretation seems to most accurately capture Kant's thinking. I will focus now on the view of rationality present in each interpretation as well as the impact this view of rationality has on the distinction between the contradiction in conception test and the contradiction in the will test.

One Contradiction or Two?

Korsgaard notes as an advantage of her Practical Contradiction Interpretation that it “should enable us to employ the same sense of contradiction in interpreting the two contradiction tests, and yet still to distinguish between them” (Korsgaard 96). Korsgaard notes that the Logical Contradiction Interpretation cannot deal with the contradiction in the will test (Korsgaard 96). Of course, it remains an open question whether the ability to deal with both tests is reason to prefer one interpretation over another.

Korsgaard is of course completely correct to say that we cannot look for a logical contradiction in maxims that involve a contradiction in the will rather than a contradiction in conception. This is the case by definition. Kant notes, of contradictions in the will, “inner impossibility is indeed not to be found” (*Gr* 424). It is for this reason that Korsgaard notes that “proponents of the Logical Contradiction Interpretation for the contradictions in conception end up with something like a utilitarian or a teleological view about contradictions in the will” (Korsgaard 96). It is no doubt true that, when it comes to contradictions in the will, no appeal can be made to a logical contradiction and another sort of contradiction must indeed be sought. However, Korsgaard takes this as a reason to abandon the Logical Contradiction Interpretation altogether. It is this step that seems questionable. It seems that Korsgaard takes it as understood that, when it comes to dealing with the distinction between contradictions in conception and contradictions in the will, we are to prefer an interpretation that can account for both sorts of contradictions. This seems to be a dubious assumption.

Korsgaard's claim, that we ought to employ one sort of interpretation to deal with both contradictions in conception and contradictions in the will, does however have an attractiveness to it. This is because a common, coherent view of rationality must be shared by both tests for contradictory maxims, while allowing for a distinction between the two tests. (This same desire for a cohesive view of rationality also fuels our interpretations of the relationship between the formulations of the categorical imperative.) However, since there are in fact two different tests for the acceptability of maxims, it is also reasonable to think that there could be two different sorts of contradictions in place. Korsgaard's claim is, however, that the gap between the contradiction in conception test and the contradiction in the will test is too large under the Logical Contradiction Interpretation. So, let us turn to examine this view.

While it is true that the Logical Contradiction Interpretation forces us to turn to another sort of interpretation for contradictions in the will, this may not mean that there is not a common underlying notion of rationality. Korsgaard notes that "proponents of the Logical Contradiction Interpretation ... end up with something like a utilitarian or teleological view about contradictions in the will" (Korsgaard 96). She goes on to suggest that it is not possible to use such a conception to deal with the contradictions in the will because it "presupposes a morality-laden conception of rationality" (Korsgaard 96). If true, this would be problematic because the categorical imperative is intended to be the principle of morality and so cannot rest on any prior suppositions about the good. However, Korsgaard's objection only holds if we think of a teleological view as one that

holds that the restrictions on what we can and cannot will come from a view of what is *good* to will. The objection does not work if we think of a teleological conception in terms of ends that are or are not rational to will. In the next chapter, I will argue that this is exactly what Kant has in mind. But first, let's look at Korsgaard's argument that the Practical Contradiction Interpretation does a better job.

Korsgaard describes the way that the Practical Contradiction Interpretation deals with the distinction between contradictions in conception and contradictions in the will in the following way:

The purpose thwarted in the case of a maxim that fails the contradiction in conception test is *the one in the maxim itself*, and so the contradiction can be said to be *in* the universalized maxim. The purpose thwarted in the case of the contradiction in the will test is not one that is in the maxim, but one that is essential to the will. (Korsgaard 97)

So, in the contradiction in conception test the purpose put forward in my maxim would be thwarted by its universalization. In the contradiction in the will test, on the other hand, there are "essential purposes" (Korsgaard 96) that all rational beings must have, "and in the world where maxims that fail these tests are universal law, these essential purposes will be thwarted, because the means of achieving them will be unavailable" (Korsgaard 96). Korsgaard goes on to give examples of these sorts of essential purposes when she says: "Examples of purposes that might be thought to be essential to the will are its general effectiveness in the pursuit of its ends, and the freedom to adopt and pursue new ends" (Korsgaard 96). I agree with Korsgaard's view that Kant may have in mind some 'essential ends'. Indeed I will try to develop this view in the next chapter as part of a

discussion of what it is that counts as rational willing. However, Korsgaard fails to notice an important difference that her interpretation would create between the thwarting of purposes involved in the contradiction in conception test and the thwarting of purposes involved in the contradiction in the will test.

Under Korsgaard's analysis, the thwarting of purposes in the contradiction in conception test is the "thwarting of the agent's *own* purpose" (Korsgaard 92). So the ends that are being frustrated are subjective. But, in the contradiction in the will test the ends being frustrated are ends that are essential to any rational agent. They are objective ends. It is Korsgaard's view regarding the contradiction in conception test that seems suspect to me. I dispute the view that Kant's test for ruling out maxims can be one that essentially involves the frustration of an agent's subjective ends. Such an interpretation seems to rest on what I believe is a mistaken view of what is entailed in rational willing, so it is to this question that we now turn.

Korsgaard on Rational Willing

Korsgaard writes that, "Kant's ethics is based on the idea that there is a specifically practical employment of reason, which is not the same as an application of theoretical reason. It includes a specifically practical kind of 'contradiction.' The argument that shows this seems to me to be an almost decisive one in favor of this [the Practical Contradiction] interpretation" (Korsgaard 93). While it is true that Kant's ethics is based on a practical employment of reason, this does not mean the "practical sense of contradiction" which Korsgaard mentions necessarily follows. Korsgaard argues for the

connection between the two by appealing to an analogy to Kant's discussion of hypothetical imperatives.

Korsgaard begins by noting Kant's view that hypothetical imperatives are analytic. She quotes the following passage from the *Groundwork*:

Whoever wills the end, so far as reason has decisive influence on his action, wills also the indispensably necessary means to it that lie in his power. This proposition, in what concerns the will, is analytical; for in willing the object as my effect, my causality as an acting cause, i.e., the use of means, is already thought, and the imperative derives the concept of necessary actions to this end from the concept of willing this end. (*Gr* 417)

It is easy to see how a violation of a hypothetical imperative would be incoherent and irrational. Elijah Millgram offers some humorous examples of this sort of irrationality such as "when you decide to go to New York but tear up the ticket that will get you there" (Millgram 527). Korsgaard then goes on to suggest: "Since this is the sort of contradiction implied in the analyticity of hypothetical imperatives, it is reasonable to think that this will be the sort of contradiction employed in the categorical imperative tests" (Korsgaard 94). Korsgaard then continues to think of the false promising example in terms of the analogue to hypothetical imperatives. She writes:

In the world of the universalized maxim, the *hypothetical* imperative from which the false promiser constructs his maxim is no longer true. It was 'If you want some ready cash, you ought to make a false promise.' But at the same time that he employs this hypothetical imperative in constructing his maxim, he wills its falsification, by willing a state of affairs (the world of the universalized maxim) in which it will be false. In that world, false promising is not a means to getting ready cash. (Korsgaard 94)

The idea here is that the sort of contradiction is analogous to one in which I will the ends but do not will the means, which would be a sort of irrational willing. The only

difference is that here I do not refuse to will the means to my proposed end, but instead will the undermining of my means when I attempt to universalize my maxim. However, Korsgaard's analysis misses an important point in the false promising example. It is not the fact that I have a particular end that makes it contradictory to make a false promise. False promises undermine themselves regardless of our motivations for making them. People do not cease to believe false promises because of the reasons they are made, but rather because they turned out to be *false*. I would not need to have *any* end in view for false promising to still be contradictory if practiced universally. (Of course I would have to will some sort of end, that my false promise be believed, in order for universal false promising to show me that the maxim contains a contradiction. However, this is not an end in the truest sense. We do not wish to be believed for its own sake but only for some further end.)

If we consider Kant's distinction between hypothetical and categorical imperatives, we might better see why Korsgaard's argument that they are analogous will not work. Marcus Singer notes that "when Kant first speaks of a categorical imperative, he says it is not conditioned by, or conditional upon, any purpose of the *agent*, and he distinguishes it from a hypothetical imperative precisely on the ground that the latter is conditional upon the purpose of the agent" (Singer 581). Singer is pointing to the fact that any categorical imperative must be able to hold regardless of an agent's purpose. This seems to be uncontroversial, but the importance of this claim will become clear in a moment. First, let's look at a few places in the *Groundwork*, where Kant makes it clear

that Singer is right to say that categorical imperatives do not take regard of the ends of the agent.

Kant begins by analyzing the concept of a good will. The good will, of course “is not good because of what it effects or accomplishes, because of its fitness to attain some proposed end, but only because of its volition, that is, it is good in itself...” (*Gr* 394). He then suggests that we ought to “explicate the concept of a will that is to be esteemed in itself and that is good apart from any further purpose.... In order to do so, we shall set before ourselves the concept of **duty**, which contains that of a good will...” (*Gr* 397). Kant of course goes on to say that only those actions done *from* duty constitute actions reflective of the good will. After this, Kant then reminds us that “an action from duty has its moral worth *not in the purpose* to be attained by it...and therefore does not depend on the realization of the object of the action but merely upon the *principle of volition* in accordance with which the action is done without regard for any object of the faculty of desire” (*Gr* 400). Finally, Kant makes the following remark *immediately* prior to setting down his categorical imperative for the first time: “Since I have deprived the will of every impulse that could arise from it from obeying some law, nothing is left but the conformity of actions as such with universal law, which alone is to serve the will as its principle...” (*Gr* 402). All of the preceding passages, familiar to any scholar of Kant’s ethical thought, point to the idea that the morality of my maxim absolutely cannot be derived from any consideration of the goals I hope to achieve through my action.

Nelson Potter, in his article, “Maxims in Kant’s Moral Philosophy,” suggests that this focus on the moral goodness of our motives is one of the elements that is left out of the “standard account” of maxims in Kant’s ethics. (Potter 62-63) He suggests that this is likely the case because many interpreters were chiefly concerned with maxims for the purpose of “finding criteria for arriving at unique correct descriptions of actions for the purposes of moral evaluation” (Potter 60).⁸ Potter’s claim that our motive is relevant seems important because Kant actually comes to the categorical imperative through such an analysis of what needs to be true in order for an action to express the good will. He does not simply derive the categorical imperative as a measure of right and wrong, independent of such a consideration of the moral worth of actions. If this is true, then it seems reasonable that any interpretation of what it is that makes proposed maxims acceptable or unacceptable must also be able to show that acceptable maxims (those that can be willed to be universal without contradiction) can also be seen as reflecting the good will. That is, acceptable maxims should be able to be seen as those done from a respect for the law as such.

The problem then with Korsgaard’s analysis is that, if my maxim is immoral if its universalization thwarts the purpose I hope to achieve by my action, it is *not* immoral if my purpose is not thwarted by such a universalization. Remember the earlier discussion of Korsgaard’s treatment of the case of the murderer at the door. Under her analysis I may lie to the murderer (and to anyone else who is lying to me and is unaware of the fact

⁸ Potter includes Marcus Singer, Onora (O’Neill) Nell, and Jonathan Harrison in this class of scholars.

that I am on to him) because my lie will be efficacious even if universally practiced.

Korsgaard's analogy between unfit maxims and hypothetical imperatives thus turns out to be an argument against her view instead of one that supports it. This is because a maxim that *can be* willed to be universal looks too much like a hypothetical imperative.

Kant's example of the trustee in *On the Common Saying: "That May Be Correct in Theory, But It Is of No Use in Practice,"* shows well that the acceptability of a maxim cannot possibly be linked to the purpose to be achieved by it. I will quote it at length:

Take for the case, for example, that someone is holding in trust something belonging to another (*depositum*), the owner of which has died, and that the owners heirs know nothing about it and can never come to know of it. We submit this case to even a child some eight or nine years old, and add that the holder of this deposit suffers at this very time (through no fault of his own) a complete reversal of his fortune and sees around him a miserable wife and children oppressed by want that he could relieve in a moment by appropriating this deposit; we add further that he is philanthropic and beneficent whereas those heirs are wealthy, hard-hearted and, besides, so thoroughly given to luxury and wastefulness that adding anything to their resources would be equivalent to throwing it into the sea. And now we ask whether, under such circumstances, it can be considered permissible for him to put this deposit to his own use. The one being questioned will undoubtedly answer, No! and, in place of any grounds, will be able to say only, *It is wrong!* – that is, it conflicts with duty. (*Theory and Practice* 286)

The most important part of the passage for our purposes is the fact that Kant makes it clear that the appropriation will never be discovered. It is important to see that undiscovered theft of goods entrusted to us could in fact be willed to be a universal practice, because given the conditions mentioned in the case, no amount of abuse will cause people to cease to use trustees in situations similar to the one mentioned in the example. Thus the agent's end of providing for his family and putting the money to

‘better use’ would never be undermined, even if universally practiced.⁹ The maxim would pass on Korsgaard’s analysis, yet Kant says that even an eight year old could see the obvious immorality of the action.¹⁰ The Logical Contradiction Interpretation, by contrast, does not suffer from this shortcoming, because it looks not at the ends of the agent but instead at the logical possibility of whether the proposed action could be conceived of as a law.

It thus seems to me that, at least in dealing with the contradiction in conception test, we ought to reject the Practical Contradiction Interpretation. My reasons for saying so rest on the above argument that the view of rationality that underlies such a contradiction seems incorrect. I would suggest instead that the Logical Contradiction Interpretation underlies that contradiction in conception test. However, as discussed earlier, Korsgaard rightly points out the fact that the Logical Contradiction Interpretation cannot deal with the contradictions in the will. This is of course true by definition, since the Logical Contradiction Interpretation is based on the impossibility of the universal performance of an action and Kant explicitly tells us that the contradiction in the will test deals with actions where “this internal impossibility is indeed not found...” (*Gr* 424).

⁹ One might argue that the efficacy of the action would be undermined if we removed some of the specifics from the case. However, Korsgaard’s analysis of the case of the murderer at the door relies on the same sort of specificity of circumstances.

¹⁰ It is worth mentioning that this example also gives us some reason to doubt authors such as Barbara Herman (see “Murder and Mayhem: Violence and Kantian Casuistry” p. 415) and Marcus Singer (see “The Categorical Imperative” p. 590) when they suggest that false-promising or lying may be acceptable actions, depending on our reasons for doing so. Herman suggests that false-promising may be permissible in order to help a friend, and Singer suggests that lying might be permissible to save a life (such as in the case of the murderer at the door). But, the trustee example shows that the person entrusted with the estate would have fairly noble reasons for keeping the money, yet, according to Kant, doing so would be clearly wrong.

We will therefore need to find a different sort of contradiction for maxims that fail the contradiction in the will test.

I will argue that there is a sort of teleological contradiction present under the contradictions in the will test and that we will be able to see that there is indeed a common view of rationality that underlies both the Logical Contradiction Interpretation, as it is applied to contradictions in conception, and the teleological contradiction I will argue for, which can be found in contradictions in the will. The common thread of rationality involves what it is that a rational agent can and can not will. A rational agent cannot will a logical contradiction (which is why maxims fail the contradiction in conception test). In addition, there are ends that a rational agent cannot will (and failure to will such ends is what causes maxims to fail the contradictions in the will test). I will defend this view in the next chapter.

CHAPTER THREE

KANT AND OBJECTIVE ENDS

If the analysis in the first chapter is correct, we seem to be, in some sense, back at our starting point. It is frequently the case that the desire to produce a certain outcome or to prevent a very bad outcome comes into conflict with ethical standards we normally hold dear. Examples such as the doctor tempted to lie on an insurance form to help a patient, or the police officer tempted to torture a suspect to save a life, help us to remember the reality of this tension. It is only fair to demand of Kant that, if we are to choose a diligent adherence to our moral duties over the attempt to promote or prevent certain ends, that this choice can be defended and that we can be shown that such a choice is ‘worth it’ somehow. I am attempting to argue that Kant indeed gives us reason to believe that such a choice is indeed ‘worth it.’ For Kant, adherence to the dictates of the moral law is valuable not only because it is the source of our dignity as human beings but also because such adherence to the moral law is seen by Kant as the only possible way to allow us to hope for a more moral world.

In order to show that Kant saw the moral law as a path towards at least the possibility of a moral world, and not as an obstacle in the way of such moral progress, we first must show that Kant saw his formalist ethics as one simultaneously concerned with ends. I will argue that Kant’s discussion of the categorical imperative and his discussion of ends that are at the same time duties are cases in which the moral law commands us to

promote, or at least forbids us from acting against, certain ends. This will help us to see that it is possible to conceive of a formalist ethics such as Kant's as having teleological concerns without also conceding to critics of Kant's that he is a closet consequentialist.

In Chapter One, I examined Christine Korsgaard's treatment of the Formula of Universal Law (or Law of Nature). I argued there that her view of what it is that causes maxims to fail Kant's tests for universalizability (the contradiction in conception test and the contradiction in the will test) blurred the line between hypothetical imperatives and categorical imperatives in a way that Kant would not admit. Korsgaard had suggested that we cannot will maxims to become universal when their universal practice would undermine our ability to accomplish the end of our action. This however turns into a question of whether it would be possible for me to accomplish my goals if my method of accomplishing those goals were to be universally practiced. Seen this way, the accomplishment of my goals (given the hypothetical universal practice of my methods for achieving them) is the determining factor for the acceptability of my maxim. This however seems to put a consideration of my subjective ends in the forefront of moral deliberation, and I have tried to show that this is not in keeping with Kant's idea of morality. However it is one thing to criticize another's approach to the idea of rational willing found in Kant's Formula of Universal Law; it is another to put forth one's own idea of the sort of view of rationality Kant had in mind. So, I intend to do so here.

My criticism of Korsgaard's interpretation of the view of rationality implied in the Formula of Universal Law was its focus on our subjective ends of action. However, this

does not mean that there are not objective ends of actions, valid for all rational beings. Yet, Kant's discussion of objective ends is, of course, controversial for several reasons. First, Kant must be able to show that rational nature actually does entail the willing of certain ends, or at least that it prohibits us from acting contrary to certain ends. This is because there must be at least one objective end, valid for all rational beings, if the categorical imperative is to have material content. Second, Kant has to avoid the trap of claiming that it is the value of these ends that make them obligatory, if he is to retain the formalist nature of his ethics.

Kant says over and over again that morality cannot be grounded in an understanding of human nature or anything else empirical. For instance, he writes:

[A]ll moral philosophy is based entirely on its pure part; and when it is applied to the human being it does not borrow the least thing from acquaintance with him (from anthropology), but gives to him, as a rational being, laws a priori.... For, that there must be such a philosophy is clear of itself from the common idea of duty and of moral laws. Everyone must grant that a law, if it is to hold morally, that is, as a ground of obligation, must carry with it absolute necessity.... (*Gr* 389)

Kant emphasizes this point several times in the *Groundwork*. The above quote appears in the preface, but Kant returns to this theme in the second section of the work. There he writes:

[W]e cannot dispute that its law is so extensive in its import that it must hold not only for human beings but for all rational beings as such, not merely under contingent conditions and with exceptions, but with absolute necessity, then it is clear that no experience could give occasion to infer even the possibility of such apodictic laws. For, by what right could we bring into unlimited respect, as a universal precept for every rational nature, what is perhaps valid only under the contingent conditions of humanity? (*Gr* 408)

He even criticizes “attempts at morality in the popular taste” (*Gr* 410) for appealing to such features of human nature as their basis. Yet, despite all of these words of caution, Kant’s derivation of the four examples of duties given to us by the categorical imperative seems to appeal to features unique to our human situation.

Three of Kant’s four examples of duties that are derived from the categorical imperative make some sort of reference to teleology. The first example, of the prohibition against a maxim of taking one’s life out of self-love seems to refer to a teleology that focuses on natural purposes. There Kant tells us that “a nature whose law it would be to destroy life itself by means of the same feeling whose destination (*Bestimmung*) is to impel toward the furtherance of life would contradict itself and would therefore not subsist as nature...” (*Gr* 422). This seems to indicate that Kant is claiming that there is some impulse within the human being that is purposively designed to “impel toward the furtherance of life.” However, a richer use of teleology appears with Kant’s third and fourth examples, of a person considering maxims of refusing to develop his or her talents and of refusing to offer aid to others, respectively. Regarding developing one’s talents, Kant writes, “he cannot possibly will that this become a universal law or be put in us as such by means of natural instinct. For, as a rational being he necessarily wills that all the capacities in him be developed, since they serve him and are given to him for all sorts of possible purposes” (*Gr* 423). Here we get explicit reference to the fact that a rational being *must* will a certain end. More accurately, Kant seems to be saying that a rational being *cannot* coherently will a certain end, since that would involve the

undermining of his or her own capacity to effect other possible ends. Kant again makes explicit reference to ends a rational being cannot coherently will when he addresses the refusal to provide aid to others. There he writes, “For a will that decided this would conflict with itself, since many cases could occur in which one would need the love and sympathy of others and in which, by such a law of nature arisen from his own will, he would rob himself of all hope of the assistance he wishes for himself” (*Gr* 423).

These three examples all *seem* then to make reference to features not of rational beings as such but instead to features unique to human beings. The first example only applies to mortal beings and further only to beings that have an inherent interest in the furtherance of their own life, regardless of the happiness contained within it. The third example makes reference to the fact that we are beings that necessarily desire the accomplishment of our ends, and the fourth example trades on the fact that we are dependent beings, who need the help of others to accomplish our ends. Still further, some commentators, such as William Frankena, claim that Kant mistakenly imagines these features to be true of human beings universally. For example, Frankena writes of the example of refusing aid to others, “it is not hard to imagine a man whose fortune is fairly sure or one who is willing to be consistent and to take the consequences of his maxim’s being universally acted on; if there are such people, Kant’s test would not suffice to establish benevolence as a duty” (Frankena, 32).

Frankena’s specific comment seems to miss the point, because it suggests that the way to see whether a rational being can will a particular maxim is to imagine whether or

not there are people who actually will that maxim and would be willing to live with it as a universal law. But this takes the empirical approach that Kant warns against. Kant, it would seem, would argue that we are not discussing what it is that people *do* will, but rather what a rational being *ought* to will. However, this does not touch on the larger criticism, that Kant is making vast assumptions about human nature when he lays out these duties.

The answer to this criticism can be found, I think, when one is able to see that Kant is trying to at once find the grounds of the moral law and also to begin to ask how such a law applies to human beings. As we will soon see, it is tempting to read him as grounding the moral law in appeals to features unique to human beings when in fact he is doing nothing of the sort. Instead, as I will argue, he attempts to ground the moral law in rational nature as such, but then also applies this moral law to human beings. The difficulty rests in the fact that Kant moves back and forth between these two tasks throughout the *Groundwork*. One of his tasks in the first two sections of the *Groundwork* is to try to give the basic features of the moral law, should it exist¹. But, Kant also often speaks of the ways the categorical imperative applies to human beings. This occurs primarily in two places. It can be seen during Kant's treatment of the question of how it is that various types of imperatives are possible, and it can also be seen during his discussions of the four examples of duties derived from the categorical imperative. In the first two sections of the *Groundwork*, Kant lays out the 'groundwork' for grounding the

¹ He does not attempt to demonstrate the actual existence of the moral law until part three of the *Groundwork*.

moral law in features of rational nature: in the form of universality, in the material of an end in itself, and in the will of an autonomous agent. Meanwhile, since these first two sections of the *Groundwork* are also concerned with distinguishing his project from popular philosophy, Kant is also occupied with other questions. These include the practical importance of finding a ground for the moral law and the application of the moral law to human beings. In order then to fully understand the way that Kant connects rationality to ends, it is important to distinguish carefully between the places in which Kant is discussing the grounding of the moral law and places in which he is attempting to show how such a law relates to the will of human beings.

The Ground vs. the Application of the Moral Law

It may sound obvious, but one cannot understate the importance of the concept of duty to Kant's ethical project. This is the case because Kant conceives duty as an action we conceive of as necessary. And of course, the action must be necessary in itself and not as a means to some other end (in which case it would only be necessary as long as we desired the other end). So, Kant's project deals with the question of whether and how it is possible for a being to conceive of an action as necessary of itself and to freely act according to such a conception. However, the project begins as largely a negative one, by which I mean that Kant initially discusses a morality based on reason alone in contrast to attempts at morality that attempt to derive their foundations based on empirical considerations.

Kant's contrast between his approach and past 'popular' approaches to morality is important because Kant is trying to show that the concept of duty, or of an action necessary in itself, is impossible if we base our notions of morality on empirical considerations. This claim is true, of course, because empirical considerations are themselves not necessary and can therefore never generate the concept of an action that is necessary in itself. I point this consideration out because it is important for us to realize that Kant's insistence that "all moral philosophy is based entirely on its pure part; and when it is applied to the human being it does not borrow the least thing from acquaintance with him...but gives to him, as a rational being, laws a priori..." (*Gr* 389), does not refer not to the application of the categorical imperative to human actions. Instead it is speaking of the grounds of moral obligation as based on a concept of a rational being, which is contrasted with attempts to base them on empirical considerations of the nature of human beings. This distinction will become important later, as it will help us understand how it is that Kant comes up with the specific duties he enumerates in the four examples he uses to illustrate the categorical imperative.

Kant has so far been discussing the need to find a ground for the moral law independent of empirical considerations. However, he soon moves on to a new task. When Kant tells us that "[o]nly a rational being has the capacity to act in *accordance with the representation* of laws, that is, in accordance with principles, or has a *will*," (*Gr* 412) he is now beginning to speak not of the possibility of the moral law but of our "capacity to act in *accordance with*" such a law. Kant is concerned with our ability to give

ourselves the law *and* to follow it, and so must be concerned with our will. He echoes this sentiment again when he writes: “Practical good, however, is that which determines the will by means of representations of reason...that is, from grounds that are valid for every rational being as such” (*Gr* 413). Here again, the grounds of the moral law must be valid for all rational beings, but Kant is not only concerned with the grounds of the moral law, but also with how one’s will can be determined to act according to this law. So again, we see a dual purpose; claims that refer to the need to ground the moral law in rational nature as such are coupled with concerns for determining a will according to such a law.

It should be clear at this point that, while Kant is discussing the fact that the moral law must be valid for all rational beings, he is also concerned with the issue of how such a law takes the form of *commands*. He notes that “the determination of such a will in conformity with objective laws is *necessitation*....The representation of an objective principle, insofar as it necessitating for a will, is called a command (of reason), and the formula of the command is called an **imperative**” (*Gr* 413). This signals a point in Kant’s discussion in which he is asking a new question: “In what ways does the moral law speak to humans as a command of reason?” And this is, of course, different from asking how the moral law is possible. This difference is in fact crucial because, while the moral law must be valid for all rational beings, imperatives are different. Kant tells us:

Hence no imperatives hold for the *divine* will and in general for a *holy* will: the “ought” is out of place here, because volition is of itself necessarily in accord with the law. Therefore imperatives are only formulae expressing the relation of

objective laws of volition in general to the subjective will of this or that rational being, for example, of the human will. (*Gr* 414)

Imperfect beings have wills that are subject to inclinations which prevent us from acting in accordance with reason, so any talk of an imperative or command only makes sense in reference to such an imperfect being. Therefore, it should not be surprising that Kant appeals to features of human existence when he applies the categorical imperative to his four examples, nor is such an application an overreaching on his part. The categorical imperative deals with the fact that “the relation of objective laws to a will that is not thoroughly good is represented as the determination of the will of a rational being through grounds of reason, indeed, but grounds to which this will is not by its nature necessarily obedient” (*Gr* 413).

Kant is concerned with the various ways in which the will can be determined by reason. Kant tells us that “every practical law represents a possible action as good and thus as necessary for a subject practically determinable by reason...” (*Gr* 414). Of course, there are various notions of ‘necessary’. On one hand there are actions that are seen as necessary only if we use them to attain some other purpose, and on the other hand there are actions that are seen as necessary because they are seen as being in themselves good (*Gr* 414). This is of course the basis for Kant’s distinction between hypothetical and categorical imperatives. In hypothetical imperatives, actions are seen as necessary because “[w]ho[m]ever wills the end also wills (insofar as reason has decisive influence on his actions) the indispensably necessary means that are within his power” (*Gr* 417). But actions are only necessary in the truest sense, that is, objectively necessary, *if* an end

is conceived of as necessary. But, these ends are all “discretionary” (*Gr* 415), that is, except one; Kant tells us that happiness is an odd exception. With happiness, the end is conceived of as one that is actual rather than contingent and as one that is necessary for human beings rather than arbitrary. Kant writes:

There is, however, one end that can be presupposed as actual in the case of all rational beings (insofar as imperatives apply to them, namely as dependent beings), and therefore one purpose that they not merely *could* have but that we can safely presuppose they all *actually do have* by a natural necessity, and that purpose is *happiness*. . . . It may be set forth not merely as necessary to some uncertain end, merely possible purpose, but to a purpose that can be presupposed surely and a priori in the case of every human being, because it belongs to his essence. (*Gr* 416)

This passage is intriguing because it may be tempting to read Kant’s contrast between hypothetical imperatives and categorical imperatives as a distinction between arbitrary and necessary ends, but this depiction is not quite complete. The problem with happiness is not that the end is arbitrary. Kant conceives it not only as actual but necessary, and as applying to all rational beings to which imperatives apply. Rather, happiness cannot serve as a ground for morality because there is no way to derive a link between the necessity of the end and the necessity of any action toward that end. As Kant tells us, “the problem of determining surely and universally which action would promote the happiness of a rational being is completely insoluble...” (*Gr* 418). This is because happiness is an indeterminate concept, so reason can’t tell us that any action is necessary as a means to the end of happiness. Additionally, although happiness is an actual end of all human beings and is thus also a natural and necessary end of *human* beings, it is not a necessary end of all rational beings. And, although the categorical imperative relates to

human beings or imperfect beings generally in its application, this imperative still must be grounded in an end valid for all rational beings as such.

So far the distinguishing feature of Kant's approach is that the categorical imperative presents an action as necessary of itself and so does not appeal to some further end. He has deprived morality of all empirical concerns and thus any concern for ends to be acquired. He tells us that "the worth of any object *to be acquired* by our action is always conditional" (*Gr* 428). So, we are left without an end for our actions, and are only left with the form of our willing. Kant writes:

[W]hen I think of a categorical imperative I know at once what it contains. For, since, the imperative contains, *beyond the law*, only the necessity of that the maxim be in conformity with this law... nothing is left with which the maxim of action is to conform but the universality of law as such; and this conformity alone is what the imperative properly represents as necessary. (*Gr* 421)

So, hypothetical imperatives present ends of action to us; most are merely possible ends, while happiness is an actual end that Kant says can be legitimately assumed in all human beings, since they are rational beings to whom imperatives apply. In contrast, the categorical imperative has been deprived of ends and we are left merely with the command that our maxims take lawfulness as their form. But, this is clearly not the complete story. *All* actions contain *some* end. We cannot use the end of happiness, or the arbitrary ends that are only subjectively determined, so Kant must go on to find a different kind of end. This end must be an objective end, that is, one valid for rational beings as such. Just as our will must take the form of universality and necessity, because it must be valid for all rational beings in order to serve as a ground of the moral law, there

must also be an end that is also necessary and given by rational nature as such. So, we turn now to Kant's discussion of this end, humanity as an end-in-itself.

Subjective Ends vs. the Objective End of Humanity as an End-In-Itself

Kant's discussion of humanity as an end-in-itself represents an important turning point in the *Groundwork*. Prior to that point Kant's discussion of duty has focused on the form of our willing, namely, the conformity to law as such. Indeed Kant spends much of the *Groundwork* prior to his discussion of humanity as an end in itself trying to steer clear of any discussion of ends because he takes great pains to separate his approach from an approach that takes empirical considerations as grounds for the moral worth of actions. Kant has stripped his moral calculus of such empirical considerations and has thus left only the form of willing as a guide for the morality of our actions. However, Kant makes it clear that maxims must not only have the form of universality but that they also must have an end (*Gr* 436). Indeed Kant indicates that without some sort of end in itself a categorical imperative would not be possible when he writes:

If then there is to be a supreme practical principle and, with respect to the human will, a categorical imperative, it must be one such that, from the representation of what is necessarily an end for everyone because it is an *end in itself*, it constitutes an *objective* principle of the will and thus can serve as a universal practical law. (*Gr* 428-29)

Since Kant has so far taken care to avoid talk of ends, and now he is claiming that an end is absolutely necessary in order for a categorical imperative to be possible, it is critical to understand the way in which Kant distinguishes this end in itself from any other end.

It is not immediately clear what sort of end could be conceived as objectively necessary, so Kant instead begins his discussion by contrasting such ends with ends of action, or subjective ends (*Gr* 428). Subjective ends, are “ends that a rational being proposes at his discretion as *effects* of his actions” (*Gr* 428), and these “are all always relative” (*Gr* 428). This is because such ends derive their worth simply from the fact that they are desired, and since desires are all contingent, such ends “can therefore furnish no universal principles, no principles valid and necessary for all rational beings...”(*Gr* 428). Kant makes it clear that, in contrast to such ends of action, “[O]bjective ends... hold for every rational being” (*Gr* 427). We thus have the primary distinction between objective and subjective ends: subjective ends are contingent or relative, whereas the objective end Kant seeks must be seen as a necessary end, valid for all rational beings. And, this distinction gives rise to a further, related, contrast between objective and subjective ends. Subjective ends, or “objects of the inclinations” (*Gr* 428) are not unconditionally valuable. Instead they “have only conditional worth” (*Gr* 428). So, an objective end cannot be an end of action or an “object *to be acquired* by our action” at all (*Gr* 428). Kant then has set himself a difficult task. He must present us with an end that is necessary and unconditionally good, but that is also an independently existing end.

Kant puts forward this independently existing end at first merely as a postulate, arguing simply that such an end in itself must exist. For, “without it nothing of *absolute worth* would be found anywhere; but if all worth were conditional and therefore contingent, then no supreme practical principle for reason could be found anywhere” (*Gr*

429). So, his first words about this end in itself are cautious. He tells us to “suppose there were something the *existence of which in itself* has an absolute worth, something which as *an end in itself* could be a ground of determinate laws; then in it, and in it alone, would lie the ground of a possible categorical imperative, that is, of a practical law” (*Gr* 428). Immediately, Kant proposes that rational nature is this end in itself. He says simply, “Now I say that the human being and in general every rational being *exists* as an end in itself...” (*Gr* 428). Kant does not provide much argument for this claim, at least not at this point. He instead only offers a few enigmatic phrases. He first notes, “rational beings are called *persons* because their nature already marks them out as an end itself” (*Gr* 428). Later he tells us that the “human being necessarily represents his own existence in this way.... But every other rational being also represents his existence in this way consequent on just the same rational ground that holds for me...” (*Gr* 429). But, even there Kant notes (in a footnote) that this assertion is only a “postulate,” to be defended later. However, his defense of this postulate is not critical for our current purposes. I wish to consider instead how Kant conceives of this end in itself, assuming it exists.

Kant has taken pains to distinguish objective ends from ends that are to be attained through our action. Kant clearly sees humanity differently. When Kant speaks of humanity as an objective end, he tells us that it is “an objective end that, whatever ends we may have, ought as law to constitute the supreme limiting condition of all subjective ends...” (*Gr* 431). He thus seems to conceive of humanity as an end that may not be

acted against, a constraint on our action, rather than an end to be promoted². This point gets made most clearly when he writes:

But since... abstraction must be made altogether from every end to be *effected* ... the end must here be thought not as an end to be effected but as an *independently existing* end, and hence thought only negatively, that is, as that which must never be acted against and which must therefore in every volition be estimated never merely as a means but always at the same time as an end. (*Gr* 437)

The matter however becomes further complicated when the possibility is introduced that humanity is not just a limiting condition, but a positive end to be realized, embodied in the duty to promote the happiness of others and to promote our own moral development (Atwell 92). Here it is not that we are trying to bring humanity into existence but rather that there is a positive duty to further it or develop it, either in our own person or that of another. And this, of course, is a source of controversy in Kant, because any talk of ends to be promoted seems to blur (if not step over) the line between subjective and objective ends.

Any duties we have must ultimately be able to be traced back to grounds that are valid for all rational beings, namely that our maxims are able to take the form of universality and that our actions have as their end an unconditioned and self-existing end.³ However, the notion of an end that is nonetheless not an end of action is one that is not immediately clear. So, we must now investigate what Kant meant by ‘humanity’. Further, we need to see how Kant connects his conception of ‘humanity’ as the end in

² Christine Korsgaard and John Atwell both come to the same conclusion. See Korsgaard’s *Creating the Kingdom of Ends*, p. 108, and Atwell’s *Ends and Principles in Kant’s Moral Thought*, chapter 5.

³ This does not, of course, mean that our duties are duties that hold for every rational being, but only that the ground of obligation is valid for rational beings as such.

itself and thus as the material ground for the moral law to two other features of his view of morality. We need to see how his view of humanity connects to his discussion of what it is rational to will, so that we can see the connection between the various formulations of the categorical imperative. Additionally, we need to see how it is that Kant connects his idea of humanity to his discussion of ends that are also duties. Once we do so we will be able to see the tie between the formalist and teleological elements of Kant's ethical thought. These connections begin to become clear when we consider Kant's four examples of duties derived from the categorical imperative.

However, I would first like to differentiate the sort of discussion I will be putting forth from a different sort of discussion of teleology in Kant. The way that Kant discusses the Formula of Universal Law has led some commentators, with Paton chief among them, to suggest that Kant sees the Formula of Universal Law as appealing to natural purposes; this is a view I do not share. In order to examine the difference between my approach and that taken by Paton, let us briefly look at his view and some criticisms of it.

Paton's View

In his essay, "Paton on the Application of the Categorical Imperative," Nelson Potter remarks that:

Paton's views concerning the application of the categorical imperative emerge in his discussion of Kant's famous four examples in relation to the "law of nature" formulation. The most novel and interesting of those views is his claim that when maxims are morally tested by being regarded as laws of nature, they are to be regarded as teleological laws. (Potter 414)

Paton indeed writes: “In every case [Kant] appeals to teleological considerations; and there is no possibility of even beginning to understand his doctrine, unless we realize that the laws of nature he has in mind are ...teleological” (Paton 149).⁴ The reason that Potter calls such a view “novel and interesting” can be seen when we take a look at what he thinks are some of the possible advantages of Paton’s view.

Potter ascribes two main benefits to Paton’s view. First, “If there are elements of teleological law involved in all moral arguments, there is no special problem about that (relatively small but important) class of arguments that involve an explicit and clear use of teleological conceptions of nature” (Potter 416). Indeed at least two of Kant’s examples of the application of the categorical imperative in the Formula of Universal Law explicitly use teleological language. In the suicide example Kant writes, “a nature whose law it would be to destroy life itself by means of the same feeling whose destination is to impel toward the furtherance of life would contradict itself and would therefore not subsist as nature” (*Gr* 422). And, in the case of developing one’s talents, Kant again uses teleological language when he writes that, “as a rational being he necessarily wills that all the capacities in him be developed, since they serve him and are given to him for all sorts of purposes” (*Gr* 423). In both examples Kant makes reference to natural purposes. These examples provide problems for commentators looking for one

⁴ Paton takes it as an assumption that, in taking our maxims as laws of nature, we must either regard them as causal laws or as teleological laws. He rejects the possibility that our maxims are to be treated as causal laws and argues that they are to instead be considered teleological laws. This argument appears on pages 148 and 149 of *The Categorical Imperative*. The argument concerning whether laws of nature are to be thought of as causal or teleological is a separate issue from an analysis of the type of teleology Paton sees as present in the Formula of the Law of Nature. It is the solely the latter issue with which I will be concerned.

cohesive interpretation of the sort of contradiction involved in willing unacceptable maxims. Whereas the case of lying, for example, can be handled well by both the Logical Contradiction Interpretation and the Practical Contradiction Interpretation which I discussed earlier, the explicit use of teleological language seems to challenge the view that either the Practical or the Logical Contradiction Interpretations can treat all of Kant's examples in a satisfactory way. If instead Paton is correct that there are teleological dimensions to all moral arguments, then we may be able to rescue our hope of finding a unified interpretation of the nature of the contradictions involved in the Formula of Universal Law.⁵

The second advantage Potter sees of Paton's approach is that "it calls attention to the importance of the *end* of action in Kant's moral philosophy" (Potter 416). Potter notes that Kant's later work shows an increasing focus on the ends of action. Of the *Metaphysics of Morals*, Potter goes so far as to say that "[t]he interpreter of Kant who reads this work realizes the importance of the conception of "objective ends" in the application of the categorical imperative" (Potter 416-417). Potter goes on to say of Paton's view:

The interpretation of the "law of nature" formulation according to which nature is to be viewed as a teleological system and the universalized maxim is to be treated as a teleological law ... seems to fulfill the need for justification for the "objective ends" Kant speaks of. Thus, given Paton's interpretation of maxims and his desire to be a defender of Kant's ethics, the "teleological law" theory of the application

⁵ Although Korsgaard raises the possibility (which she rejects) that there may be one type of contradiction involved in the examples using perfect duties and another in the examples involving imperfect duties, it is interesting to note that Kant's language favors a different division. Kant uses explicitly teleological language in the two cases that involve duties to oneself, while it is absent from the examples of duties to others. I will return to this consideration later.

of the categorical imperative may seem to be the most satisfactory one available.
(Potter 417)

Potter is on to something here. A thorough discussion of the categorical imperative must be able to show connections to Kant's discussions of objective ends, and teleological considerations are best suited to such a task. However, this does not mean that Paton's approach, of asking us to treat our universalized maxims as teleological laws, is correct. Both Potter and Korsgaard reject Paton's view, and I believe that they are right to do so. Yet, as I hope to show a bit later on, while Paton's treatment may be wrong, we may still be able to find that teleology nonetheless plays an important role in the application of the categorical imperative. Let us now take a closer look at Paton's view and some reasons for rejecting it.

Paton focuses his discussion around his view that "on Kant's view to conceive human nature as governed by teleological law is to suppose a complete harmony of ends both within the race and within the individual" (Paton 150). He then goes on to explain how this view would get fleshed out in the application of the categorical imperative.

Paton writes:

We can consider human nature *as if* there were such a systematic harmony of ends in accordance with a law of nature; and we can ask whether any proposed maxim, if it were made a law of nature, would fit into such a systematic harmony. Some maxims would destroy such a systematic harmony, while others would merely fail to foster it, and this seems to be the basis of the distinction between perfect and imperfect duties. (Paton 150)

One might ask how it is that Paton feels so certain of the claim that we ought to think of human nature as governed by teleological law, to be thought of in terms of a harmony of ends.

In his attempt to show that the “laws of nature” Kant has in mind when we attempt to universalize a maxim are teleological in nature, Paton notes that both action and human nature are essentially purposive and thus that conceiving of maxims (which necessarily refer to actions) as laws of (human) nature entails teleology (Paton 151).

However, Potter points out the fact that there seems to be a flaw in Paton’s thinking here.

Potter writes:

One of Paton’s arguments is as follows: He notes, as we have, that for Kant actions are essentially purposive and that all maxims mention an end. He then directly concludes that, therefore, a maxim when regarded as a law of nature is a teleological law. But this is a *non sequitur*. If Paton thinks that the conclusion follows directly, then it seems that he is confusing “conscious purposes” and “natural purposes”. It is the latter that are involved in teleological laws; it is the former that are involved in actions. (Potter 420)

Potter’s criticism appears to be a valid one. It is not necessarily the case that because we see ourselves as aiming at particular ends that human actions do indeed aim at systematic harmony of purposes.

Yet Christine Korsgaard finds reason to defend Paton, at least on this point.

While Paton may be wrong to leap from purposiveness to teleological laws, Korsgaard points out that there is indeed evidence in Kant that reason does aim at a harmony of purposes. She suggests that Paton could have indeed cited comments in the *Critique of Pure Reason*, in which Kant specifically says that the highest formal unity of reason is in

fact a harmony of purposes, to support his view (Korsgaard 91). If then reason aims at a harmony of purposes and a rational agent as such must be committed to such a harmony of purposes, “we may take it that when a maxim universalized as a law of nature could lead to such a harmony of human purpose, that maxim is fit to be adopted also as a moral law” (Paton 151). Thus for Paton, “the order of nature is a type for the systematic harmony of human purposes” (Korsgaard 91).

However, Korsgaard goes on to point out that, while it might be true that “Kant thinks rationality commits us to a harmony of purposes.... this conclusion does not necessarily imply that when we reason morally we reason *from* such a harmony” (Korsgaard 91). Korsgaard’s point is an important one. There is much in Kant that suggests that rationality and indeed morality ultimately aim for a harmony of purposes. (One thinks of Kant’s treatment of the highest good as an example of this.) However, this does not mean that we can imagine what such a unity of purposes would look like and reason back toward a derivation of duties. The only way this would be possible would be if we could “somehow establish that the proposed natural purpose of the action type is one needed for the systematic harmony of purposes and therefore is one the agent must will” (Korsgaard 91).

If we turn to the issue of actually attempting to see how various actions contribute to the systematic harmony of ends, both Potter and Korsgaard are able to see several obvious problems with Paton’s view. First, as Potter points out, there is no empirical evidence that there are any teleological laws at all, and if there were, there is no way of

knowing what they are (Potter 417). However, it is clear upon reading Paton's analysis that he does not claim that there actually are teleological laws of nature but only that, for Kant, we are to act *as if* there were.⁶ But Potter is willing to grant Paton as much and goes on to consider the implications. Potter notes of the approach of acting 'as if' there were teleological laws of nature: "there are a good many problems with any such conception: (i) It is not clear how one ascertains whether a given proposed "as if" principle holds, pertains, or is "valid" or not" (Potter 418). Potter goes on to say that this problem leads to another issue. It is possible for us to come up with many possible ideas concerning the possible purpose(s) of a supposed action type. He uses the example of sexual organs, which could be alternatively conceived of as having the purpose of pleasure or of reproduction (Potter 418). Potter's criticism is correct, as far as it goes. We cannot know the natural purpose of organs (even if there were one). However, Christine Korsgaard makes it clear that the view Potter is criticizing is not identical with Paton's actual view.

Potter's example of sexual organs is not ideal, for it confuses what Christine Korsgaard calls the 'simple view' of teleology with a more complex view, which she sees in Paton (Korsgaard 87). Korsgaard describes the 'simple view' of the sort of contradiction involved in a failure to universalize a maxim as "usually understood in this way: the contradiction emerges when an action or instinct is used in a way that is inconsistent with its natural purpose, or is not used in a way that its natural purpose calls

⁶ Paton actually puts the phrase 'as if' in italics himself to stress this point. (See both p. 149 and p. 150 of *The Categorical Imperative*)

for” (Korsgaard 87). Potter is thus criticizing the ‘simple view’ when he talks about our inability to really know what that natural purpose of a given action, instinct, or organ. But, as we have already seen, Paton is treating teleological laws as aiming not at natural purposes but a systematic harmony of purposes. The confusion is created because Kant’s own language, in the examples of suicide and developing one’s talents refer to natural purposes. Further, as Korsgaard points out, Paton refers to this language to support his own interpretation (Korsgaard 88). It is when we turn to Paton’s treatment of the examples of providing assistance to others and of promising to repay loans that we get a better sense of his view that actions and practices aim at a systematic harmony of purposes.

Paton takes the view that, in treating our maxims as teleological laws, “we assume empirical knowledge of nature (particularly of human nature)” (Paton 146-147). Paton takes the view that Kant must have had such assumptions about human nature in mind in his treatment of his four examples. Of the case of the duty to provide assistance to those in need, Paton claims: “The argument turns on the fact that human beings are in need of mutual help, and that only by means of mutual help can the systematic harmony of purposes be attained” (Paton 152). In support of this interpretation Paton refers to a passage in the *Metaphysics of Morals* that reads:

Consequently the maxim of common interest, of beneficence toward those in need, is a universal duty of human beings, just because they are to be considered fellow human beings, that is, rational beings with needs, united by nature in one dwelling place so that they can help one another. (MM 453)

The above passage supports the first half of Paton's claim that the argument turns on the fact that human beings are beings in need of mutual help, but it makes no reference to the second part of his claim, that the purpose of this mutual aid is to promote a systematic harmony of ends. It is at least as plausible to argue that, as a being in need of aid, I cannot will the non-existence of aid (in the form of my refusal to provide it for others) while I *must*, as a rational agent in need of such aid (a human being), simultaneously will the existence of such aid.

The view that the purpose of a given practice must be to promote a systematic harmony of purposes is more apparent in Paton's treatment of the example of the promise to repay loans. However, Paton is on even less firm ground here. Paton claims that "Kant assumes empirical knowledge that the purpose of such promises is to produce trust and so to get out of financial difficulties" (Paton 152). He also goes on to say:

What Kant says is true enough so far as it goes, but it does not offer a satisfactory basis for moral judgment unless we make the further assumption that the keeping of such promises and the mutual confidence thereby aroused are essential factors in the systematic harmony of human purposes. (Paton 153)

The problem with this is to be found in the assumption Paton tells us we must make. It is difficult to see how one would show that a systematic harmony of purposes would be rendered impossible without the 'mutual confidence' produced by keeping promises.

Korsgaard makes the same point when she writes:

[T]he idea that a rational being is committed to a harmony of purposes will only help us with the Formula of Universal Law if we can somehow establish that the proposed natural purpose of the action-type is one needed for the systematic harmony of all human purposes and therefore is one that the agent must will. (Korsgaard 91)

It is important to make clear that I am not denying that morality aims at a systematic harmony of purposes. What does seem implausible however is the view that we can simply assume that we know beforehand which actions promote this systematic harmony and that we can then use this knowledge to test our maxims. Paton seems to have the equation backwards. We first need to be able to find out what purposes a rational agent must will. It may be that a rational agent must will various objective ends and that among these we should include the promotion of a systematic harmony of purposes. It may even be that the other objective ends a rational agent must will all help to promote such a systematic harmony. However, we cannot simply imagine that they do, assume we know how it is that they do so, and then use this as our basis for ascertaining specific duties. We must first find the necessary ends of a rational agent, and only then can we ask how it is that they might further the systematic harmony of human purposes.

In order to see how Kant ties his understanding of rational willing to a consideration of ends we must either not act against or must promote, we need to return to a careful examination of the four examples he gives of duties derived from the categorical imperative. The examples of duties derived from the categorical imperative collectively are examples of what it is that either cannot be rationally conceived or rationally willed without contradiction. I would suggest that, rather than using Kant's first formulation of the categorical imperative to make sense of these examples (the difficulty of which we saw in chapter one), we work in the other direction, using the

examples to make sense of both the first and second formulations of the categorical imperative. This in turn will also help us see the connection between these two formulae.

Kant's Four Examples and the Connection between the Formula of Universal Law and the Formula of Humanity

Kant's first two examples are supposedly of examples of maxims which cannot even be thought without contradiction. The example of the lying promise seems fairly straightforward whether we approach it from the Formula of Universal Law or the Formula of Humanity. The Formula of Universal Law tells us: "*act only in accordance with that maxim through which you can at the same time will that it become a universal law*" (*Gr* 421). Kant uses this to test the maxim: "when I believe myself to be in need of money I shall borrow money and promise to repay it, even though I know that this will never happen" (*Gr* 422). Kant says that such a maxim fails because, "the universality of a law that everyone, when he believes himself in need, could promise whatever he pleases with the intention of not keeping it would make the promise and the end one might have in it impossible..." (*Gr* 422). While it is true that the contradiction Kant has in mind may be taken in various ways,⁷ one way to take this is simply that it is a flat logical contradiction to both will some thing and will its opposite simultaneously. In willing the universal adoption of lying promises a person on one hand wills that people make lying promises. (Of course, 'people' here refers simply to him or herself). On the other hand this same person wills that people (everyone else) do not make lying promises, since if they did, "no one would believe what was promised him and would

⁷ See the discussion in chapter one.

laugh at all such expressions as vain pretenses” (*Gr* 422). A rational being cannot coherently will (or even conceive) a logical contradiction.

Turning to the Formula of Humanity, this same example again seems to be the easiest to handle. Here Kant says, “he who has it in mind to make a false promise to others sees at once that he wants to make use of another human being *merely as a means...*” (*Gr* 430). So far, the problem with the false promise is that, in making such a promise, we simply fail to realize that persons, as “rational beings...are always to be valued at the same time as ends” (*Gr* 430). Here Kant is relying simply on the notion that rational beings as ends in themselves are unconditionally valuable, which, as we have seen above, is one of the ways he differentiates the end in itself from subjective ends. So, of course, when I use a person for my own benefit I clearly treat them as only conditionally valuable and as something whose value can be subordinated to some other end. And, Kant has just finished arguing that, if the moral law is to have legitimacy at all, there must be an unconditionally valuable end. Yet, Kant’s language in this example also lets us see an additional reason why the false promise treats a person merely as a means, and this begins to help us see how exactly Kant conceives of the notion of humanity as an end in itself.

Kant tells us that the person to whom we would lie cannot “cannot possibly agree to my way of behaving toward him, and so himself contain the end of this action” (*Gr* 430). Kant’s language here can be confusing, since it seems obvious that a person can in fact contain the end on behalf of which the lie was told. For example, a doctor might lie

to a patient in order to make sure the patient stays on a proper course of treatment, and the patient of course could approve of the end the doctor has in mind: the patient's health. So, Kant must mean something else when he says that the person to whom we would lie cannot possibly agree to our behavior. Any person to whom we would lie cannot "contain the end of the action" because the lie denies the possibility that he or she could choose the end we have chosen. Here we get our first hint of Kant's view of exactly what it is that comprises humanity. Humanity begins to emerge as the ability to set ends for one's self. This is why, when Kant tells us that rational beings are to be valued at the same time as ends, he finishes the sentence by saying, "that is, only as beings who must also be able to contain in themselves the end of the very same action" (*Gr* 430). Kant even includes a footnote in which he takes pains to separate his notion of treating persons as ends from the idea that we ought to treat others as we would have them treat us, so as to avoid just the sort of interpretation depicted in the above example of the doctor and patient.

So, both formulae of the categorical imperative can be used in a relatively straightforward way to derive the conclusion that we may not make a false promise. However, if we constrain ourselves to this example alone, there is much that is lacking. First, the connection between the two justifications of this duty, given by the two formulations, is not clear. And, since this is the case, there is nothing in the derivation of the duty to refrain from making lying promises under the Formula of Humanity that helps to better understand the notion of rational willing present in the derivation of the same

duty under the Formula of Universal Law. Instead, we must look at Kant's other three examples to better see the relationship between rationality and ends.

The suicide example is particularly frustrating when approached from the perspective of the Formula of Universal Law. This example seems radically different from the lying promise example in that a person could very well will both that he or she end his or her own life when continuing to live "threatens more trouble than it promises agreeableness..." (*Gr* 422), and also allow that the maxim holds for all others as well. So, we don't seem to have the same sort of logical contradiction that is present in the example of false promising. Additionally, Christine Korsgaard tells us frankly that her "Practical Contradiction Interpretation" of the Formula of Universal Law cannot handle this example either. She tells us, "On the Practical Contradiction Interpretation we cannot get an analysis of that case, for the suicide's purpose, if it is release from his own misery, will not be thwarted by universal practice" (Korsgaard 100). Indeed it is the overtly teleological language in this example that could best support Paton's view that Kant is making use of an appeal to natural purposes.⁸ However, I have already argued that Paton's view must ultimately be rejected. This then seems to leave us without a clear understanding of how it is that a maxim of suicide is contradictory under Kant's conception of rationality. Kant's treatment of the same duty from the perspective of the Formula of Humanity takes a more helpful approach.

⁸ Kant's defense of the duty to refrain from suicide under the Formula of Universal Law runs as follows: "a nature whose law it would be to destroy life itself by means of the same feeling whose destination (Bestimmung) is to impel toward the furtherance of life would contradict itself and would therefore not subsist as nature..." (*Gr* 422).

The suicide example is in some sense the clearest example of the command that we ought always treat humanity as an end in itself. Kant tells us that, if a person “destroys himself in order to escape from a trying condition he makes use of a person *merely as a means* to maintain a tolerable condition up to the end of life” (*Gr* 430). Since the very concept of the end in itself is that humanity cannot be made to be subservient to *any* other end, even happiness, then suicide is clearly impermissible. We can however supplement this interpretation in light of the understanding of humanity we have taken from the example of the lying promise. The examples are clearly very different in some sense. In the case of the false promise we know that the person to whom we would lie cannot “himself contain the end of this action” (*Gr* 430) because the lie denies the possibility of acting on full information and thus of freely choosing one’s own end. We have no such problem in the suicide example, which seems to be an end we can freely choose. Yet, if humanity consists of the ability to set ends for one self, as the lying promise example suggests it does, then suicide is impermissible because it cuts off this possibility permanently. In choosing suicide we give up our ability to set ends, just as we would if we voluntarily gave ourselves up to slavery. This analysis allows us to begin to see a connection between the Formula of Humanity and what Kant sees as rational willing. If the false promise example shows us that it is irrational to will a logical contradiction, then the suicide example as seen from the point of view of the Formula of Humanity offers us the possibility that it is also irrational to will that we give up our

ability to set ends. In order to see if this is correct, we shall now consider the last two of Kant's examples.

Kant's final two examples present challenges for us when they are approached from the perspective of the Formula of Humanity. This is because Kant has already told us that humanity as an end in itself is to be thought only negatively, as a limit of what we may do to our selves and others, and that it is not an end to be attained. Yet, these two examples seem to present cases where humanity is conceived as something to be promoted rather than as something we cannot act against. The problem that arises is that, if there are ends that it is our duty to promote, Kant must establish the grounds of these duties without lapsing into consequentialism. We must show that the duties to be beneficent and to develop one's own talents can be linked to Kant's notion of what is entailed by his notion of rational willing. This is because, if there are in fact duties to promote certain ends, it is vital to show that their status as duties is independent from any value they have.

In order to show the continuity between the Formula of Universal Law and the Formula of Humanity, we will also need to try to show that the duties of beneficence and of developing one's talents do not present us with a notion of humanity as something that ought to be promoted. This is the approach taken by John Atwell. I agree with Atwell that the duties of beneficence and of developing one's talents can be seen in a way so that they do not necessarily imply a view of humanity itself as an end to be promoted (and I will discuss this view briefly). But, I would also suggest that these duties can be

grounded in Kant's idea of rational willing. This is important because Kant will go on to develop his discussion of these duties in his discussion of ends that are also duties in the *Doctrine of Virtue*, and there the question of what they imply about humanity is not enough to address the issue of how the duties themselves are derived. We begin with Atwell's approach.

Atwell's approach is to suggest that, "Kant never espoused a 'positive' conception of the objective end (humanity, man, or persons) in the first place" (Atwell 92). He points out the fact that one can have positive duties without having a 'positive' notion of the end that grounds said duties. Atwell writes, "There is surely no inconsistency in claiming that (1) the end-in-itself can be conceived only negatively, and (2) our actions can agree or harmonize with the end-in-itself both negatively and positively" (Atwell 94). Atwell gives the example of freedom as an analogy. He notes that, "human freedom can be conceived only negatively (e.g., as the absence of hindrances) yet conduct regarding human freedom can 'agree' with it negatively (e.g., by never creating hindrances for yourself or others) as well as positively (e.g., by removing hindrances)" (Atwell 94). He even goes so far as to point out that we could call the removing of hindrances as a 'promotion' of freedom in the sense it is helpful, but this does not change the fact that freedom is just the absence of hindrances and so is not something that is an end to be attained (Atwell 94-95). Christine Korsgaard also takes this approach. When she discusses the case of the duty to develop one's talents, as addressed from the perspective of the Formula of Humanity, she writes, "This, indeed, is as close as Kant comes to

assigning a positive function to humanity as an end. What makes this possible is the fact that rational nature is a sort of capacity....not an end to be effected...” (Korsgaard 127). If Kant indeed conceives of humanity as the capacity to set ends, then both Atwell and Korsgaard’s reasoning is sound. And as the first two examples have already suggested, this is exactly what Kant conceives of as humanity. Later, in the *Doctrine of Virtue*, Kant clearly identifies humanity this way when he writes, “The capacity to set oneself an end – any end whatsoever – is what characterizes humanity (as distinguished from animality)...” (*MM* 392).

Kant’s language also suggests that the duties of beneficence and of developing one’s talents are duties that agree with humanity, and he stays away from language that would instead imply that in these duties we see a view of humanity as itself an end to be achieved. Regarding the duty to develop one’s talents Kant tells us that “it is not enough that the action does not conflict with humanity in our person as an end in itself; it must also *harmonize with it*” (*Gr* 430). Regarding the duty of beneficence Kant uses similar language. He notes that “there is still only a negative and not a positive agreement with humanity as an end in itself unless everyone also tries...to further the ends of others” (*Gr* 430). Here again, Kant is careful to use the phrase ‘positive agreement with,’ which is in keeping with Atwell and Korsgaard’s interpretation. So, Atwell and Korsgaard are correct to say that the existence of positive duties does not imply a positive conception of humanity, since, as we have already seen, humanity is a self-existing end.

If humanity is indeed to be thought as the ability to set ends, this understanding helps us to make sense of Kant's third and fourth examples, the duties to develop one's talents and to be beneficent respectively. In fact, without this understanding of humanity, Kant's discussion of these examples, at least from the perspective of the Formula of Humanity, would be incomplete. Kant says very little about each of these duties when discussing the Formula of Humanity. Of the duty to develop our talents Kant's only guidance comes in the passage cited above, in which he tells us that our actions must do more than avoid conflicts with humanity but must also have a positive agreement with humanity as an end in itself (*Gr* 430). But this does little to tell us either about what it is that Kant sees as comprising humanity or about how this specific duty can be derived. The same holds true with the duty of beneficence, where Kant again primarily confines himself to the reminder that beneficence is a duty implied by a positive agreement with humanity as an end.

However, Kant does offer us one important additional insight in his discussion here of the duty of beneficence. He tells us that "the natural end that all human beings have is their own happiness" (*Gr* 430) and that "the ends of a subject who is an end in himself must as far as possible also be my ends..." (*Gr* 430). This passage does give us some insight into the notion of humanity Kant has in mind. Earlier I suggested, following others, that Kant's idea of humanity is that it consists in our capacity to set ends. We might refine this a little to say make it clear that to set an end is to will the attainment of that end, and thus the means to its attainment. So beneficence seems to be

required of us because, to deny a person of the means to happiness is to deny them of means to an end that is a necessary end of human beings. This is important because it cannot be the case that we act against humanity as an end in itself every time we thwart any end whatsoever. Instead, we violate humanity when we do one of two things. Such violations occur when we undermine a person's ability to set and realize ends in general, or when we deprive persons of the means to an end that is considered a necessary end of human beings, and one such end is of course happiness. However, in order to develop our understanding further, we should turn to Kant's treatment of these same examples under the Formula of Universal Law.

Kant's examples of the duties to develop one's talents and to be beneficent are the examples that present the most difficulty in interpretation. Not only do these examples tempt us to misinterpret Kant as espousing a positive conception of humanity as an end to be promoted or attained, they also seem to be the two examples that make appeals to contingent features of human existence rather than to features of rationality as such. But both of these apparent problems can easily be overcome. We have already seen that it is possible to conceive of something only negatively but to also suggest that there are ways that our duties can "harmonize" with such a conception positively and that Kant's treatment of the third and fourth examples from the perspective of the Formula of Humanity makes just such a distinction. Regarding the accusation that Kant appeals to features of the human condition and not to features of rationality per se when he is deriving his examples of duties, I have already argued that, although it is true that Kant

does make reference to features specific to the human condition, there is nothing problematic with his doing so. This is because, as Barbara Herman suggests:

When Kant speaks of excluding empirical considerations from morality he has two related things in mind. First, that the foundation of morality is to be non-empirical....Second, that each of us is subject to moral requirements independent of any contingent, empirical ends we happen to have....Kant does not need to argue (nor does he) that the content of morality is to be determined without regard to the empirical nature of things. (Herman 590)

Earlier in this chapter I have tried to flesh out the point here made by Herman. I have tried to show the distinction between grounding the moral law and its application and I have also tried to give a brief summary of Kant's contrast between subjective and objective ends. However, we still need to return to an examination of Kant's third and fourth examples, as seen from the perspective of the Formula of Universal Law, because we need to show convincingly that these duties are derived from a coherent view of what it is or is not rational to will. In doing so, we should simply remember that we are asking what it is rational or irrational for a *dependent and imperfect* yet rational being to will.

When discussing the duty to develop one's talents, as seen from the perspective of the Formula of Universal Law, Kant tells us that we have such a duty because a person cannot will the maxim of refusing such development. Kant writes that this is so because, "as a rational being he necessarily wills that all the capacities in him be developed, since they serve him and are given to him for all sorts of possible purposes" (*Gr* 423). The claim here seems to be that it is irrational to undermine one's capacity not to set ends but to realize them. All rational beings set ends, and the willing of an end entails the willing of the necessary means to that end. This in no way is tantamount to Kant claiming that

we ought to will *certain* ends. We simply cannot deprive ourselves of the means necessary to attain whatever ends we do set. Since our talents are one of the tools we have at our disposal that assist us in the realization of our ends, we cannot will a maxim of ridding ourselves of such a tool. Herman makes the point this way when she writes that “the ends that are necessary to sustain oneself as a rational being cannot (on rational grounds) be given up. Insofar as one has ends at all, one has already willed the continued exercise of one’s agency as a rational being” (Herman 586). Willing the maxim of refusing to develop one’s talents is irrational because it entails willing a maxim of undermining our very “agency as a rational being.” This same point is made even more powerfully in the example of the duty to be beneficent.

When Kant presents his claim that there is a duty to be beneficent for the first time, under the Formula of Universal Law, his language is admittedly such that it opens Kant up easily to criticism. Kant says that we cannot will a maxim of refusing aid to others because:

For a will that decided this would conflict with itself, since many cases could occur in which one would need the love and sympathy of others and in which, by such a law of nature arisen from his own will, he would rob himself of all hope of the assistance he wishes for himself. (*Gr* 423)

This passage lends itself to a view of the Formula of Universal Law such as the one espoused by Korsgaard, and against which I argued in Chapter One, that the reason a maxim of non-beneficence cannot be willed without contradiction is that such a maxim, when universalized, would undercut my own particular goals. But, as Herman reminds us: “To center moral deliberation on a strategy for even hypothetical self-protection

provides a lesson one would not have expected Kant to endorse” (Herman 582).⁹ So, Kant cannot be appealing to a claim that we will want the help of others and so, out of self-interest, ought to extend the same help to others that we would expect for ourselves.

We need to instead realize that rational dependent beings, by definition, need the help of others to attain our various ends, and so cannot will that we forgo such aid. Herman puts the matter this way: “The argument to a contradiction in the will asks, in effect, whether it could be rational for a human being to renounce irrevocably the resource (means) of the help of others” (Herman 585). We can see that indeed we cannot renounce such means because we would be willing to make the attainment of certain ends impossible. We cannot know which ends would or would not be possible without the aid of others, but we can be sure that there will indeed be at least some ends that are impossible without such aid. This is true simply because we are imperfect and dependent beings, with “natural limitations of our powers as agents” (Herman 587). The difference between this approach and Korsgaard’s is indeed a subtle one but it might be summarized the following way. A given person can indeed be content to refuse aid from others, so such a person could indeed will a maxim of non-beneficence if his or her only concern was whether or not the kindness would be repaid. In such a case, the concern is with the aid itself. However, since the lack of aid also deprives us of the ability to attain what Herman calls our “true needs,” (Herman 587) what we cannot rationally will is the abandonment of the means to the attainment of ends we cannot achieve on our own.

⁹ In “Mutual Aid and Respect for Persons” Herman gives an excellent account of the ways that a view such as Korsgaard’s opens Kant up to various casuistry concerns.

Here the concern is not with whether or not we *want* assistance, but with the fact that, as dependent beings, we, by definition, *need it*.

We can now see that, in all four of the examples of duties derived from the categorical imperative Kant avoids two possible problems. It was important to be able to show that Kant is able to clearly link all four of these duties to a coherent view of rationality. This became possible once we began to understand humanity as the capacity to set and realize our ends and to use this conception to help us make sense of Kant's derivations of the four examples from both the perspective of the Formula of Humanity and the Formula of Universal Law. Second, since Kant has to make sure to separate humanity, as an end in itself, from subjective ends, we also needed to show that Kant does not treat humanity as an end to be achieved but rather as a limiting condition on our subjective ends. Understanding humanity as the capacity to set and attain ends, allows us to see how it is that, in the four examples, Kant can consistently conceive of humanity as an end that is not to be acted against rather than as one that ought to be promoted.

Ends that Are At the Same Time Duties

We have so far established that Kant is able to derive, from both the perspective of the Formula of Universal Law and the Formula of Humanity, a series of ends we cannot rationally will (and in some cases cannot even conceive). But Kant's discussion becomes more complicated when he begins to deal with ends that are at the same time duties. Even if humanity is not an end to be promoted, we still need to determine whether he claims there are such duties. We must turn now to Kant's discussion of ends

that are at the same time duties that appears in the *Metaphysics of Morals*. It is there that the true defense of how it is possible that we can have ends that are duties is given. In the *Groundwork* this is left un-argued for.

Kant reaffirms his view that a discussion of ends cannot take us to a concept of duty, while insisting that the concept of duty will take us to a discussion of ends. He explains: “For since there are free actions there must also be ends to which, as their objects, these actions are directed” (MM 385). But, Kant quickly adds that these ends must include among them ends that are at the same time duties. He tells us: “For were there no such ends, then all ends would hold for practical reason only as a means to other ends; and since there can be no action without an end, a categorical imperative would be impossible” (MM 385). This is of course because the categorical imperative tells us that we must perform certain actions simply because it is our duty to do so, so there must be ends that we are commanded to value for their own sake. Otherwise, all ends would be contingent and all commands only hypothetical.

So far, Kant has not stepped beyond his discussion of humanity as an end in itself, since humanity is an end that is objectively necessary. However, Kant goes further at this point and begins to discuss more specific ends that he claims duty commands us to promote. The two chief ends he mentions are our own perfection and the happiness of others, which have already been embodied in the last two examples of duties that were depicted in the *Groundwork*. Once Kant claims that there are ends we have a duty to *promote* he again risks the criticism that such ends are by definition subjective ends, and

not objective ends commanded by duty. Earlier, we noted that the existence of these two sets of ends (self-perfection and beneficence) did not entail the notion that humanity was itself an end to be promoted, but this does not necessarily mean that they are themselves not such ends. But here again, John Atwell takes the view that the happiness of others and our own perfection are not to be seen as ends to be promoted.

Atwell suggests that all ends of action are by definition subjective ends. He asks: “Is everything a person attempts to bring about, by the very fact that he tries to effect it, a subjective end, i.e., something having only a relative value, thus something serving only in hypothetical imperatives?” (Atwell 90). He then goes on to answer his own question: “To answer affirmatively, as my criticism suggests we should, would seem to make a categorical imperative impossible” (Atwell 90). Atwell goes on to offer a solution to this difficulty, arguing that ends that are at the same time duties should not be thought of as ends of action at all. He writes that “*no objective end* (not even ‘an end which is a duty’) *is ever an end of action*- understanding ‘an end of action’ as that which the action is devised to bring about” (Atwell 90). To support his claim Atwell turns to Kant’s distinction between objective ends and subjective ends. He notes that, “Whatever anyone intends to effect by means of an action is always a subjective end and never an objective end...” (Atwell 90). In support, he cites Kant’s claim that “the value of all objects to be produced by our action is always conditioned” (*Gr* 428). By contrast, “The value of an objective end, on the other hand, is absolute and unconditioned, hence no objective end can be the end of an action, i.e., an ‘object’ that can ‘be produced’ by our action” (Atwell

90-91). The question then is whether one's own perfection and the happiness of others can be seen as ends to be effected by action and thus, according to Atwell, subjective ends.

It is safe to say that our own perfection is not an end of action. Kant presents us with two different kinds of self-perfection: natural perfection and moral perfection. Kant describes the former this way: "Natural perfection is the cultivation of any capacities whatever for furthering ends set forth by reason" (*MM* 392). This kind of perfection seems to go hand in hand with a respect for our own humanity as the capacity to set ends. Kant clarifies this point when he continues to say:

The capacity to set oneself an end – any end whatsoever – is what characterizes humanity (as distinguished from animality)...In other words, the human being has a duty to cultivate the crude predisposition of his nature, by which the animal is first raised into the human being. It is therefore a duty in itself. (*MM* 392)

This first kind of perfection thus involves simply acting in ways that do not interfere with our ability to set ends, as well as acting in ways that enhance our ability to set ends, but there are no set ends established as ones we ought to effect. The second kind of self-perfection Kant presents as a duty is the duty of moral self-perfection. Kant tells us that, "The greatest perfection of a human being is to do his duty from duty (for the law to be not only the rule but also the incentive of his actions)" (*MM* 392). As Atwell suggests, "Moral self perfection is not an 'extrinsic' end that one can attempt to produce by means of action..." (Atwell 97). This is because such moral perfection consists in acting from a certain disposition, namely the good will, and the good will is not something external to us that we bring about but rather is an element of our willing itself. Thus, this end is the

kind of self-existing end that Kant espoused as his notion of humanity. The happiness of others, however, seems more likely to be an end we can bring about by our action, and so seems more likely to appear to be a subjective end.

There are really two ways to argue that happiness is not a subjective end of action. The first is to suggest that we cannot possibly be the causal agent of someone else's happiness, since Kant understands happiness to be the fulfillment of all of one's aims. The happiness of others would then be at best an end we can contribute to via our actions, not an end of those actions themselves. The second avenue is to argue, as Atwell does, that when Kant tells us that we have a duty of beneficence that he is not arguing that we have a duty to create happiness in others but rather that we cannot possibly will a maxim of refusal to provide aid to others (Atwell 98). Atwell's reasoning does seem to be in keeping with Kant's own articulation of the reason why beneficence is an end that is also a duty. Kant presents the following argument:

The reason that it is a duty to be beneficent is this: since our self-love cannot be separated from our need to be loved (helped in the case of need) by others as well, we therefore make ourselves an end for others; and the only way this maxim can be binding is through its qualification as a universal law, hence through our will to make others our ends as well. The happiness of others is therefore an end that is also a duty. (*MM* 393)

Kant's statement points back to the analysis we made earlier. The duty of beneficence stems from the fact that a rational, dependent being cannot coherently will the deprivation of means necessary to the attainment of one's various ends, so we cannot deny this same resource to others. Kant's discussion of ends that are also duties therefore seems in

keeping with our earlier analyses; he establishes these ends on grounds that are not separate from the formulations of the categorical imperative.

Now that we can say with confidence that Kant is on firm ground when he argues for the validity of the existence of ends that are at the same time duties, it may be worthwhile to look more in depth at one such kind of duty. I have in mind the duties to oneself in terms of the duties regarding our own perfection, especially our moral perfection. This is important because I want to argue that Kant's commitment to the rational elements of his ethics, namely universality, humanity as an end in itself, and a self-legislating or autonomous will, translate to a commitment to moral perfection of ourselves and a striving toward the moral perfection of the world as a whole. I will treat the former here and the latter in the next chapter.

The issue of duties to oneself is an interesting one in Kant's ethics. In a utilitarian approach to morality, any moral obligation or moral responsibility to one's self only exists because of its larger utilitarian benefit. In an Aristotelian kind of virtue ethic, by contrast, the concern is on the development of one's self with the implication being that a moral concern for others will be the natural result of a sufficiently virtuous character, but the idea of *obligation* to others is absent. Kant's approach is to show that the ground of our obligations, both to our selves and to others, in fact stems from the ability to put our selves under obligation. So, in a sense, all obligations for Kant are obligations to our selves. This is because the ground of morality in Kant's approach is the autonomy of practical reason. Kant makes this point when he writes that "I can recognize that I am

under obligation to others only insofar as I at the same time put myself under obligation, since the law by virtue of which I regard myself as being under obligation proceeds in every case from my own practical reason; and in being constrained by my own reason, I am also the one constraining myself” (*MM* 418). This view, that obligation or duty can only make sense as seen as emanating from an autonomous will takes us back to the *Groundwork* and to the link Kant makes there between the categorical imperative and autonomy.

In the *Groundwork* Kant makes it clear that “all practical lawgiving” (*Gr* 431) can be seen in the form of universality and the end of humanity as an end in itself (*Gr* 431). But he goes on to say that “from this there follows now the third practical principle of the will, as supreme condition of its harmony with practical reason, the idea *of the will of every rational being as a will giving universal law*” (*Gr* 431). Our humanity in our own person is that which enables us to give this moral law, and, in order for it to be a binding law, it must take the form of universality. Together we get the notion of a legislating or autonomous will. Kant makes it clear that his discussion of imperatives, both from the perspective of their form and from the perspective of their ends, had been so far “only *assumed* to be categorical because we had to make such an assumption if we wanted to explain the concept of duty” (*Gr* 431). The concept of autonomy is necessary in order to show that commands can in fact be categorical. This is “just because of the idea of giving universal law *it is based on no interest* and therefore, among all possible imperatives, can alone be *unconditional...*” (*Gr* 432). The ability to hold oneself under

obligation is thus the ultimate ground of moral obligation; without it, the universality of the categorical imperative becomes a mere assumption. Likewise humanity as an end in itself, according to Kant, only makes sense if reason is autonomous. He writes that “the will of a rational being must always be regarded as at the same time *lawgiving*, since otherwise it could not be thought of as an *end in itself*” (*Gr* 434). The important thing to stress here is that one’s ability to act as a lawgiver stands as the ultimate condition for the categorical imperative while also allowing us to see a bridge, together with the notion of humanity as an end in itself, to the more teleological elements of Kant’s ethics.

In the *Doctrine of Virtue* Kant reminds us that freedom consists, oddly enough, in being under obligation, specifically obligation to ourselves. He writes that “the ... human being thought in terms of his personality, that is, as a being endowed with inner freedom...is regarded as a being that can be put under obligation, and indeed under obligation to himself (to the humanity in his own person)” (*MM* 418).¹⁰ He goes even further than this, saying that the duties to our selves are the ground of all our duties, saying: “For suppose there were no such duties [to our selves]: then there would be no duties whatsoever, and no external duties either” (*MM* 418).

Kant divides duties to our selves into negative and positive duties. “Negative duties forbid a human being to act contrary to the **end** of his nature and so have merely to do with his moral *self-preservation*...” (*MM* 419). In contrast, “positive duties, which

¹⁰ Kant seems here to be using ‘humanity’ in a way that is different from the way he uses it when he goes through his four examples under the Formula of Humanity of the categorical imperative. There ‘humanity’ seems to be the capacity to set ends. Here however, this capacity seems to be referred to as ‘personality,’ while ‘humanity’ is reserved for the aspect of ourselves that is law giving; it represents our autonomous will.

command him to make a certain object of choice his end, concern the perfecting of himself” (*MM* 419). So, while Atwell may be right that one’s perfection is not itself an end to be promoted, this does not mean that it does not entail that we “make a certain object of choice our end.” Atwell’s approach of claiming that the ends that are duties are not ends to be promoted is a helpful one. However, the more important issue is whether any end that is a duty can ultimately be grounded in appeal to Kant’s understanding of rational nature as an end in itself that can both give and be obligated by universal laws.

Kant ultimately does claim that there are positive duties to ourselves that go beyond duties to refrain from depriving ourselves of the ability to set ends. We must also actively cultivate our ability to set these ends. Kant tells us that, “A human being has a duty to himself to cultivate his natural powers...as means to all sorts of possible ends. – He owes it to himself (as a rational being) not to leave idle and, as it were, rusting away the natural predispositions and capacities that his reason can someday use...”(*MM* 445). This claim indicates that we are, so to speak, works in progress. Humanity, the ability to set ends, is a capacity that can be dampened or heightened, and it is our duty to develop this capacity as far as it is in our power to do so. Kant makes it clear that we cannot necessarily attain moral perfection, but that we nonetheless have a duty to strive toward it. He writes that, “It is a human being’s duty to strive for this [moral] perfection, but not to reach it (in this life), and his compliance with this duty can, accordingly consist only in continual progress” (*MM* 446). This claim will become important in Kant’s discussion of the highest good, and I will take it up in more depth in the next chapter. For now, it is

sufficient that we are able to see that Kant's notion of a rational being, which can both give and follow the universal law, and which has itself, as the ability to set ends, as its own end, implies positive duties to one's self. (Kant will also go on to discuss duties to others, but I will take these up in the next chapter.)

The purpose of this chapter was to defend Kant's discussion of ends by showing the legitimacy of his account of objective ends and to show how such ends can be linked to Kant's notion of rationality. There are three important elements to Kant's notion of rationality: rational willing as taking the form of universality, and rational nature both as an end in itself and as consisting in the capacity to be obligated to one's self as a lawgiver. I have tried to show how all three of these elements are present in Kant's discussions of duties, both as they appear in the *Groundwork* and as they reappear in the *Metaphysics of Morals*. There is more that can be said on this subject, but my goal here has simply been to provide an argument for the legitimacy of Kant's discussion of ends. This discussion is a sort of 'bridge' to the discussion that is upcoming in the next chapters.

If Kant's ethics is to demand so much of us, prohibiting us from even lying to the murderer at the door, then we need to be able to show that this rigorism is indeed worth the price, and I believe it is possible to defend Kant's view that it is indeed worthwhile. But this defense is only possible if Kant's view of morality can be seen as being concerned with how it is that we can bring about a more perfect world. And of course, this means that we must first show that Kant can be shown to be on solid grounds in his

discussion of ends in general, which has been the goal of the present chapter. In the next chapter I will turn to Kant's discussion of our duty to bring about a more perfect world, which includes his discussions of the Kingdom of Ends and the Highest Good. This discussion will also consider the place Kant's views of nature, God, and hope play in Kant's vision. Then, in the last chapter, I will argue that Kant's approach is the *only* approach that at least entitles us to hope for and believe in the possibility of a morally perfect world.

CHAPTER FOUR

THE MORAL WORLD, MORAL PROGRESS, AND THE MORAL LAW

This dissertation is ultimately guided by two questions. First, is Kant's rigorous formalism really as rigorous as it seems? In other words, is it really the case that Kant should be taken at his word when he claims, for example, that we cannot lie, even to a murderer at the door? In the first chapter I argued that, despite possible wishes to the contrary, we must indeed answer this first question in the affirmative. Kant's conception of the categorical imperative indeed involves a sort of inflexibility that is either the greatest strength or greatest weakness of his ethics. I tried to show this by arguing, against Christine Korsgaard, that the various formulations of the categorical imperative should be read as leading to the same conclusions in moral decision making. This discussion inevitably leads to the second question: Can the strict dictates of Kant's ethics be defended when they seem to require us to allow great harms to occur?

In Chapter Two I argued that the beginning of the answer to this question rests with understanding that Kant's ethics, while quite strict, is not entirely formalistic. I argued that we can see a strong element of teleology present even in the way Kant understands the categorical imperative. Further, I argued that the teleological elements of the categorical imperative lead to a better understanding of Kant's discussion of ends that are also duties. These ends that are also duties chiefly surround the duty to cultivate our talents and the duty to promote the happiness of others, and I have argued that these ends

that are also duties can be derived from both the Formula of Humanity and the Formula of Universal Law. I attempted to demonstrate this by offering an explanation and defense of Kant's claim that one cannot rationally will either a maxim of refusing to give aid to others or a maxim of refusing to cultivate one's talents. In short, a rational being cannot coherently deprive one's self or another person the means necessary for realizing various ends, nor can one refuse to develop the capacity for setting and realizing ends. This also helped to show the connection between the Formula of Universal Law and the Formula of Humanity, since 'humanity' is this very capacity to set and realize ends.

In the ends that are also duties, we have the view that we have a duty to *promote*, or at the very least to refrain from interfering with, the ability to set and realize ends, both in our selves and others. This discussion leads to Kant's discussion of the kingdom of ends, since, as we will soon see, such a kingdom is one in which all citizens are able to freely set and realize their various ends. It also takes us to the heart of the argument in this dissertation: adherence to the moral law is valuable not just because rationality is valuable for its own sake, but because it can be seen as helping to lead to a more moral world. Kant makes this point when he says, "Morality consists, then, in the reference of all action to the lawgiving by which alone a kingdom of ends is possible" (*Gr* 434). Morality is thus, on Kant's account, bound up with the project of creating a 'kingdom of ends.' If this is true, then the commands of duty can be seen as helping to make a kingdom of ends possible. Thus, if it is only through such adherence to the moral law that

the kingdom of ends is possible, then adherence to the dictates of morality also paves the path for a better world. This is the next part of the answer to those who would criticize Kant for his allegiance to the strict dictates of the moral law when such allegiance seems to make it difficult to confront evils such as that depicted in the case of the murderer at the door. Those who use the murderer at the door as an example of what is wrong with Kant's ethics portray it as an example of a choice between a Kantian morality and the desire to prevent evil or to promote a better world. However, if adherence to the moral law in fact is seen by Kant himself as bound up with creating a 'kingdom of ends,' which is one term Kant uses for a moral world¹, then this choice may be a false one, and following the moral law and promoting a moral world can be seen as two equally important parts of Kant's ethical project.

In order to better see how Kant connects the Formula of the Kingdom of Ends to the other formulations of the categorical imperative, it is helpful to consider the way he connects these formulations through the concept of autonomy. Kant begins by linking

¹ In the *Groundwork* Kant notes that "a world of rational beings (*mundus intelligibilis*) as a kingdom of ends is possible, through the giving of their own laws by all persons as members" (*Gr* 438). In the first *Critique* we see a link between the kingdom of ends and the highest good when Kant writes: "Now in an intelligible world, that is, in the moral world...such a system, in which happiness is bound up with and proportioned to morality, can be conceived as necessary..., since rational beings, under the guidance of such principles, would themselves be the authors both of their own enduring well-being and that of others" (*CR* A 810/B 838). The connection between happiness and morality runs through most of Kant's discussions of the highest good, which I will examine in depth within this chapter. We see the connection to the kingdom of ends when we note the recurrence of Kant's phrase 'intelligible world' or '*mundus intelligibilis*'. But Kant's discussion of the moral world is not confined to the concepts of the kingdom of ends and the highest good in the world. For example, he also uses terms such as the 'ethical community' in the *Religion* (*Rel* 99). This chapter will largely be devoted to tracing the links between these concepts while arguing that they share the common feature of a world of rational beings living under common freedom and that Kant understood this condition to be possible in the world of sense.

both the Formula of Universal Law and the Formula of Humanity to autonomy. He writes:

[T]he ground of all practical lawgiving lies...*objectively in the rule* and the form of universality which makes it fit to be a law...; *subjectively*, however, it lies in the *end*; but the subject of all ends is every rational being as an end in itself...; from this there follows now the third practical principle of the will...the idea of *the will of every rational being as a will giving universal law*. (Gr 431)

The idea of every rational being *giving* the law to itself is important because it, like the other two conditions of a command of practical reason, the form of universality and an objective end, is a necessary condition in order for the law to be unconditional. Kant makes this point when he says that “the *principle* of every human will as *a will giving universal law through all its maxims*...would be very *well suited* to be the categorical imperative by this: that just because of the idea of giving universal law *it is based on no interest* and therefore, among all possible imperatives, can alone be *unconditional*...” (Gr 432). Having connected the principle of autonomy to the previous two formulations of the categorical imperative, Kant then goes on to connect it to the Formula of the Kingdom of Ends.

Kant claims that “the concept of every rational being as one who must regard himself as giving universal law through all the maxims of his will...leads to a very fruitful concept dependent upon it, namely that *of a kingdom of ends*” (Gr 433). Kant then continues on to say, “[b]y a *kingdom* I understand a systematic union of various rational beings through common laws” (Gr 433). These laws are laws rational beings give to themselves and follow, and so we see the connection between the kingdom of ends and the principle of autonomy and the Formula of Universal Law. But these laws

also present an objective end to practical reason, so Kant also connects the kingdom of ends to the Formula of Humanity when he writes:

For, all rational beings stand under the *law* that each of them is to treat himself and all others *never merely as means* but always *at the same time as ends in themselves*. But from this there arises a systematic union of rational beings through common objective laws, that is, a kingdom, which can be called a kingdom of ends (admittedly only an ideal) because what these laws have as their purpose is just the relation of these beings to one another as ends and means. (*Gr* 433)

A kingdom of ends is thus an ideal world in which every member of the kingdom treats him or her self and others as ends in themselves. And further, we are also to serve in part as means to the ends of others, wherein each member of the kingdom of ends has a duty to help promote the free ends of others. Thus, it seems Allen Wood is correct when he notes, in *Kant's Ethical Thought*, that the Formula of the Kingdom of Ends “is derived by combining ideas drawn from all the previous formulas of the moral law” (Wood 166). We see this in the element of adhering to the law, which we also give to ourselves, as well as the command to treat others as ends in themselves. Another way to put this is to say that the kingdom of ends is thus the condition that would result if we were to treat humanity as an end in itself or if we were to adopt only those maxims that can be willed to hold universally without contradiction. It is this claim, that there is a state of affairs that will result if everyone were to follow the categorical imperative, which is the essential addition that comes with the Formula of the Kingdom of Ends.

Kant tells us that each formulation of the categorical imperative represents a progression from the formulations that have preceded it. This point is made most clearly

when Kant discusses the various formulations of the categorical imperative as a way of explaining the various components of maxims. Kant writes:

All maxims have, namely, 1) *a form*, which consists in universality.... 2) *a matter*, namely an end.... 3) *a complete determination* of all maxims by means of that formula, namely that all maxims from one's own lawgiving are to harmonize with a possible kingdom of ends as with a kingdom of nature.... A progression takes place here, as through the categories of the *unity* of the form of the will (its universality), the *plurality* of the matter (of objects, i.e., of ends), and the *allness* or totality of the system of these. (*Gr* 437)

This is important because I wish to show that Kant's system is not just compatible with ends, but ultimately is fundamentally concerned with ends in the sense that our moral action ultimately must be seen as helping to promote a more moral world. Many commentators over the last few decades have made compelling cases that Kant's ethical thought is ultimately concerned with ends, especially the end of a more moral world², but what I hope to show in this chapter is not only the extent to which Kant was concerned with promoting a moral world, but also that he sees such promotion as inseparable from adherence to the moral law.

Kant alludes to both of these concerns when he writes that "such a kingdom of ends would actually come into existence through maxims whose rule the categorical imperative prescribes to all rational beings *if they were universally followed*" (*Gr* 438). Here he seems to indicate not only that the kingdom of ends could be thought of as a moral world possible in the world of sense but also that it is only possible on the condition that everyone were to work in accordance with moral laws. Kant goes on to

² See for example, Keith Ward's "Kant's Teleological Ethics". He makes the case that Kant saw his ethical project as tied to both moral perfection and the promotion of a "harmonious community" (Ward 247-8). Jennifer Moore also makes the case that Kant's objective ends are almost entirely "interpersonal or social in character" (Moore 54).

further connect these two dimensions of the kingdom of ends. First he writes that “[m]orality consists, then, in the reference of all action to the lawgiving by which alone a kingdom of ends is possible” (*Gr* 434). Second, Kant makes it clear that we must however always regard ourselves as lawgiving members of such a kingdom (*Gr* 434). Taken together, these two claims are extremely significant: we must act as though we are already in the kingdom of ends, acting as lawgivers in such a kingdom, and further, it is only through such actions that a kingdom of ends can be possible at all. Our duty is thus to make the kingdom of ends possible, but we can only do this by acting as law makers in such a kingdom. The connection between these two claims is absolutely essential for any understanding of how it is that Kant thinks about the relationship between the moral law and moral progress. Any attempt to bring about the kingdom of ends by means that conflict with the moral law would be obviously anathema to Kant’s approach to ethics. The two notions, of moral progress and adherence to the moral law, are inseparable from one another. It is this claim that allows us to better understand and defend Kant’s strict adherence to duty, even in the face of apparent injustices: it is only through such an adherence to morality that a kingdom of ends is possible.

Allen Wood shows us how this would look in practice by looking at Kant’s example of the duty of beneficence. He writes, “if I make another’s happiness my end, I thereby bring about a harmony or coincidence of ends, and to this extent contribute to a realm of ends, whereas to the extent that I refuse to make the happiness of others my end I prevent a realm of ends” (Wood 168). Wood then goes on to summarize the basic tone of the Formula of the Kingdom of Ends well when he writes that the formula “seems to

require...the *exclusion* of ends that *in principle* cannot be shared between rational beings (such as those requiring deception or coercion) and the *furthering* of ends that unite people (such as those involving mutual respect and mutual aid)” (Wood 169). What the concept of the kingdom of ends thus introduces is an element of cosmopolitanism that is not present in the earlier conceptions of the categorical imperative.³ However, in order to understand Kant’s conception of the moral world more fully and to see how it is connected to the moral law, much still needs to be done. First, we must examine what such a world looks like. Second, we must look at the extent to which Kant thought such a world was possible, and at the question of what is necessary in order for it to be brought about.

Kant’s Conception of a Moral World and its Possibility in the World of Sense

Kant’s discussions of a moral world extend through many of his writings and the notion of a moral world can be seen both in his discussions of the kingdom of ends and his discussions of the highest good. Kant remarks in the First *Critique*: “I entitle the world a *moral world*, in so far as it may be in accordance with all moral laws...” (A 808/B 836). Thus, for Jennifer Moore the moral world is, “at a minimum ... virtually indistinguishable from the kingdom of ends...” (Moore 55). Such a world, according to Kant, is one in which “we leave out of account all the hindrances to morality” (A 810 / B 838). Kant then goes on to describe the moral world further, linking it to happiness and

³ Works such as Jennifer Moore’s “Kant’s Ethical Community”, Robert Louden’s “Kant’s Impure Ethics” and Victoria Wike’s “Kant on Happiness in Ethics” explore this issue in some depth, emphasizing the fact that the moral world is one in which not only is the goal a social one but that additionally, it is one that rational agents must work together to bring about. While I discuss this to some extent later in the chapter, it is not the main focus of my argument here. I am primarily concerned with the claim that such a world is made possible only through adherence to the moral law.

morality, when he says “freedom, partly inspired by and partly restricted by moral laws, would itself be the cause of general happiness, since rational beings, under the guidance of such principles, would themselves be the authors both of their own enduring well-being and of that of others” (A 810/B 838).⁴ In the Third *Critique* Kant clearly thinks of the highest good in terms of this moral world as happiness conditioned by morality.⁵ For example, Kant writes that “the highest physical good that is possible in the world and which can be promoted, as far as it is up to us, as a final end, is happiness – under the objective condition of the concordance of humans with the law of morality, as the worthiness to be happy” (CJ 450). For Kant then, the moral world can be seen as a world in which there is general happiness as conditioned by general morality, as a product of unfettered free action.

Defining the moral world in this way allows us to see the ways that Kant connects his notion of the moral world to his discussions of autonomy and to his discussions of objective ends, or ends that are at the same time duties. Our analysis in Chapter Two

⁴ What complicates things here is that, in Kant’s descriptions of this moral world, which he comes to associate with the highest good, he frequently makes use of the notion of happiness proportioned to morality. In the *Critique of Pure Reason*, between A 810/ B 838 and A 815 / B 843 Kant uses phrasing that indicates either happiness proportional to morality or in exact proportion to worthiness to be happy several times. Several authors, most notably Pauline Kleingeld, Andrews Reath, and Victoria Wike, have made the case that there are indeed separate yet related concepts of the highest good present within Kant’s thought. See Kleingeld’s “What Can the Virtuous Hope For? Re-reading Kant’s Doctrine of the Highest Good,” Reath’s “Two Conceptions of the Highest Good in Kant,” and Wike’s *Kant on Happiness in Ethics*, especially chapters 5 and 6. The interpretation of the exact relationship between these two notions varies from author to author. While Reath draws the distinction as one between a secular notion and a theological notion, Wike articulates the distinction more in terms of an individual goal and a social goal (see especially p. 128 of Wike). Wike, Kleingeld and Reath also each make the point that the highest good as a social goal is a more important dimension of Kant’s conception of the highest good than his ‘theological’ notion or notion of the highest good as an individual goal. I am in agreement with these thinkers on this point and take this point up later in the chapter.

⁵ As I will discuss later, this should not be confused with happiness *proportional* to virtue.

showed that Kant is on firm footing when he claims that we have a duty to cultivate our talents and we have at least the minimal duty towards others of beneficence. Now, we can see the link between these two duties and the moral world. The duty to cultivate talents and the inability to will a maxim of non-beneficence both point to the idea that a rational being must will, both for one's self and for others, the maximal possibility of realizing one's freely chosen ends. The moral world is the world in which all freely chosen ends (that are compatible with morality) are able to be realized.

In the *Religion* we see a similar conception of the moral world, which Kant refers to in terms of an "ethical community" or "church visible". Kant writes:

An ethical community under divine legislation is a *church* which, inasmuch as it is not the object of a possible experience, is called the *church invisible* (the mere idea of the union of all upright human beings under direct yet moral divine world-governance, as serves for the archetype of any such governance to be founded by human beings). The *church visible* is the actual union of human beings into a whole that accords with this ideal... The true (visible) church is one that displays the (moral) kingdom of God on earth inasmuch as the latter can be realized through human beings. (*Rel* 101)

Two features of the above passage are important to note. First, the kingdom of God on earth involves a 'union of human beings into a whole' and so involves a notion of ethical community⁶ or moral world, and second, Kant notes here that this moral world is 'on

⁶ Kant uses this phrase frequently within Part III of the *Religion*. For example, Kant writes that "since the duties of virtue concern the entire human race, the concept of an ethical community always refers to the ideal of a totality of human beings..." (*Rel* 96). Kant goes on to say that "an ethical community is conceivable only as a people under divine commands ..., and indeed *in accordance with the laws of virtue* (*Rel* 99). He also links this concept to the highest good when he says that "this highest moral good will not be brought about solely through the striving of one individual for his own moral perfection but requires rather a union of such persons into a whole toward that very end..." (*Rel* 98). He does not use the phrase 'ethical community' in the passage, but it is clearly implied, as the title of the sections is "The Human Being Ought to Leave the Ethical State of Nature in Order to Become a Member of an Ethical Community".

earth' and is a state of affairs that 'can be realized through human beings.' This second point suggests that Kant saw the moral world as one possible in the world of sense.

Kant, however, gives some contradictory indicators when it comes to the question of whether or not the moral world is a world possible in the world of sense or whether it is simply an ideal. For example, he says that such a world is "only an idea, the carrying out of which rests on the condition that *everyone* does what he ought..." (A 810 / B 838), yet he also says "we must assume that moral world to be a consequence of our conduct in the world of sense..." (A 811 / B 839). I would argue that Kant did in fact see the moral world as a possibility in the world of sense. To see that this is true, we must look at the extent to which Kant emphasized the possibility of the moral world in his writings. Then we will examine the grounds we have for assuming that such a world can be seen as being able to be possibly brought into existence in the world of sense.

It must be noted that a discussion of the possibility of the moral world can be taken to mean two very different things. On the one hand, we could simply be referring to the theoretical possibility of the moral world, such that the moral world can be shown to be a valid idea of reason. On the other hand, we can also discuss the practical possibility of the existence of a moral world, in which the moral world can in fact be brought into existence in the world of sense. In what follows, it will be important for us to see that Kant sees the moral world, as the highest good in the world, as possible in both these senses.

One way that we can think of the moral world is as the final end of reason. This is, of course, one of the roles Kant ascribes to the highest good. Kant makes it clear that

the highest good, as the final end of reason, must be at least theoretically possible. He writes that pure practical reason “seeks the unconditioned totality of the object of pure practical reason, under the name the *highest good*” (*CPR* 108). He then reminds us of the connection between the moral law and the highest good by emphasizing the fact that the highest good is not the determining ground of morality; the moral law is. “Hence, though the highest good may be the whole *object* of a pure practical reason ... it is not on that account to be taken as its *determining ground*, and the moral law alone must be viewed as the ground for making the highest good and its realization or promotion the object” (*CPR* 109). Moral action thus aims at the objective end of the highest good, but the desire to create the highest good cannot and should not be the way we determine what is in fact moral.⁷ Further, this end, the highest good, is only possible as an object of action through the moral law. But, Kant reminds us that the fact that the moral law is the determining ground of morality does not mean that the highest good is therefore an inessential concept. He tells us that the:

highest good is not merely *object*: the concept of it and the representation of its existence as possible by our practical reason are at the same time the *determining ground* of the pure will because in that case the moral law, already included and thought in this concept, and no other object, in fact determines the will in accordance with the principle of autonomy. (*CPR* 110)

What Kant is saying here is that the highest good, as the unconditioned end of practical reason, contains the concept of the moral law within it, and so is a determining ground of

⁷ Klaus Dusing echoes the claim that the highest good is the ultimate end or moral action when he writes that “das höchste Gut entspringt erst auf der Anwendung der reinen Moral auf die Teleologie eines endlichen Willens; es ist nicht schon vorgegebene, sondern allererst zu entwerfende höchste Zweck unseres sittlichen Tuns in der Welt” (Dusing 33). The claim that highest good is critical as an end of action stems from the fact that a finite and free will must set ends, and these ends must have a highest end. Thus, according to Dusing, “Ein endlicher und zugleich freier Wille muss sich ja Zwecke setzen, wenn er überhaupt in seiner Welt wirken will” (Dusing 33).

the will in that it determines the will, along with the moral law, since it provides the necessary end of the moral law. Kant goes even further in drawing the connection between the moral law and the highest good when he says: “If the highest good is impossible in accordance with practical rules, then the moral law, which commands us to promote it, must be fantastic and directed toward empty imaginary ends and must therefore in itself be false” (*CPR* 114). We thus have an interesting relationship between the moral law and the highest good. The possibility of the highest good is conceptually necessary in order to provide the moral law with its necessary end, without which the entire validity of the moral law would be drawn into question.

However, all that the above argument has shown is that the highest good, as the final end of reason, must be *conceived* of as possible, and not necessarily that is possible in the actual world. Kant must do more than this if he is going to talk about a moral world that can in fact be brought about by human agency. Andrews Reath makes the point that “it would make no sense to talk about willing an action directed at an object...unless it were something that we could imagine as a result of human agency. The ...point, then, is that Kant’s definition of the good is that it should apply to possible human ends” (Reath 597). The fact that Kant makes it clear that we do in fact have a duty to promote the highest good also seems to lead us to the conclusion that the highest good must be possible through our actions, since we cannot have a duty to the impossible.⁸ In order to better appreciate the fact that Kant took the view that the moral

⁸ One of the places in which Kant makes this point is in the Dialectic of the *Critique of Practical Reason*, during his discussions of the postulate of the immortality of the soul and the postulate of the existence of God. In his discussion of the postulate of the immortality of the soul Kant notes that “*complete conformity*

world was indeed a human possibility, we must look more closely at his discussion of promoting such a world.

In the *Religion* Kant notes that, “every species of rational beings is objectively – in the idea of reason – destined to a common end, namely the promotion of the highest good as a good common to all” (*Rel* 97-98). We also see Kant’s articulation of the highest good as a social goal in the Third *Critique* when he talks of the “**highest good** to be achieved in the world through freedom ... commanded by practical pure reason for the best possible realization of that end, and which thus must be assumed to be possible” (*CJ* 469). Here the key phrases are “in the world” and “through freedom”. These phrases remind us that the highest good being envisioned here is one that exists in the world of sense and is brought about through human agency. This description appears multiple times within the Third *Critique*. For example, Kant writes: “We are determined *a priori* by reason to promote with all our powers what is best in the world, which consists in the combination of the greatest good for rational beings in the world with the highest condition of the good for them, i.e., the combination of universal happiness with the most lawful morality” (*CPJ* 453). The duty Kant is speaking of here is a duty to bring about a world in which both happiness and morality are able to be realized in all people to the fullest extent possible. Evidence of this can be seen in the phrase “universal happiness”.

of dispositions with the moral law as the supreme condition of the highest good. This conformity must therefore be just as possible as its object is, since it is contained in the same command to promote the object” (*CPR* 122). Then, in the postulate of the existence of God, he writes that “it was a duty for us to promote the highest good; hence there is in us not merely the warrant but also the necessity...to presuppose the possibility of this highest good...” (*CPR* 125). In both places, it is implicit that we cannot have a duty to the impossible, since he claims that the fact that we have a duty forces us to presuppose the possibility of the object of that duty.

In the Second *Critique* we also see a description of the highest good as a state of affairs to be brought about by human agency and as one in which all rational beings are to share. We can see this, for example, when Kant discusses the ectypal world (*CPR* 44). Kant notes that “the moral law in fact transfers us, in idea, into a nature in which pure reason, if it were accompanied with suitable physical power, would produce the highest good, and it determines our will to confer on the sensible world the form of a whole of rational beings” (*CPR* 44). The phrase “if it were accompanied by suitable physical power” is important because it seems to point to the limits of human agency in bringing about the highest good conceived of as a moral world.

Kant has argued that we have a duty to bring a moral world into existence. And, if there is such a duty, it must be possible to bring such a world into existence. It is clear that one person cannot create a moral world on one’s own, and Kant frequently reminds us that our duty is to promote, not to attain the highest good.⁹ While we can say that any individual only has a duty to promote the highest good, this would still mean that the human race as a whole would have a duty to bring it about.¹⁰ Kant indicates this when he

⁹ John Silber and Gerald Barnes both make strong cases that our duty is to promote the highest good. See Barnes’ “A Defense of Kant’s Conception of the Highest Good” and Silber’s “Kant’s Conception of the Highest Good as Immanent and Transcendent” in which he argues that the highest good is a regulative obligation in that its *attainment* is a goal to strive towards, while its *promotion* is a constitutive obligation, which informs our actual duties.

¹⁰ In fact, as Klaus Dusing points out, the very fact that one individual cannot bring about the highest good gives us reason to think of the highest good as a social goal brought about through mutual effort. Dusing makes the point when he writes, “Das höchste Gut eines Einzelnen ist nur innerhalb des höchsten Gutes aller vernünftigen und freien Wesen möglich.” (Dusing 18). This passage also points to the claim that the highest good of an individual and the highest good as a social goal are indeed different, but related concepts, as the highest good of one, according to Dusing, cannot be achieved without the highest good of all.

notes that we have a special sort of duty to create the highest good in the form of an ethical community. He writes,

Now here we have a duty *sui generis*, not of human beings toward human beings but of the human race toward itself. For every species of rational beings is objectively – in the idea of reason – destined to a common end, namely the promotion of the highest good as a good common to all. (*Rel* 97)

However, even as an entire human race, it seems that human powers alone can only *promote* a moral world; we also need to assume that nature can be conceived in a way so that human efforts to bring about the moral world will not be in vain.¹¹ Kant's writings on history all indicate an awareness of this issue as they take up the issue of progress toward a moral world while emphasizing the aspects of this moral progress that are out of human hands.

Nature and History - Progress toward the Highest Good

Kant's view that nature, particularly through its role in history, can be seen as working toward moral progress is important because it helps us see that Kant sees the highest good as more than a conceptual 'placeholder' for the final end of reason and that it is actually a state of affairs that can be made possible. This is important for two reasons. First, it underscores the degree to which Kant is concerned with the concept of

¹¹ As will become clear later in the chapter, the conformity of nature to making the world hospitable to moral action depends on the belief in a moral author of nature, God. In the final section of this chapter I argue that the highest good as a moral world possible in the world of sense exists along side another notion of the highest good, which focuses on proportioning happiness to virtue *within the individual*. This latter notion of the highest good dominates the Dialectic of the *Critique of Practical Reason*. While immortality is a necessary postulate for the realization of that conception of the highest good, it is not necessary for the possibility of the highest good as a moral world possible in a world of sense. In fact, I would suggest that moral progress through history serves an analogous role to the role provided by the postulate of the immortality of the soul. In both cases, full virtue is unlikely (if not impossible) in an individual life, so an indefinite progress is postulated. In the case of a moral world, the progress takes places in history; in the case of the individual, the afterlife is a necessary postulate.

progress toward a more moral world throughout his ethical writings. Second, it helps to refute arguments that would suggest that we can have no duty to promote the highest good, since it is not possible to realize it.

Keith Ward, referring to works such as the third *Critique* and Kant's writings on history, notes that in these works, "teleology is clearly subordinated to a higher purpose, without which there is no value in nature whatsoever" (Ward 248). Ward notes the connection between this trend and the issue of the possibility of the highest good. He writes:

[T]he paradox of Kant's ethics is that, though moral action is thus essentially concerned with the fulfillment of natural and moral perfection, such fulfillment is impossible in this world....For such a paradoxical view of morality to be intelligible at all, one must postulate that, somehow nature can be determined that moral effort will not finally be in vain. (Ward 258)

David Lindstedt also argues that a careful look at Kant's works on history will provide us with sufficient reason to believe that "Kant seems desperately concerned that the world not be dismissed as purposeless, and the possibility of the highest good on earth be secured" (Lindstedt 129). Keith Ward goes further, noting not only that Kant wishes to see nature as serving the purpose of helping to make the highest good possible, but that indeed Kant sees it as working this way. He notes that the "historical destiny of nature is thus to form a being which must be fashioned through discord and strife to transcend its animal condition and realize its essential superiority to nature, through its moral freedom" (Ward 248). In order to see if these claims are true, we turn now to our own look at some of Kant's works on history.

Even a cursory examination of *Idea for a Universal History from a Cosmopolitan Point of View* shows us at once that Kant thought of nature as unfolding through history an unseen path toward moral progress. Take for example the following passage from the beginning of the treatise:

[H]istory, which is concerned with narrating these appearances [human actions], permits us to hope that if we attend to the play of freedom of the human will in the large, we may be able to discern a regular movement in it, and that what seems complex and chaotic in the single individual may be seen from the standpoint of the human race as a whole to be a steady and progressive though slow evolution of its original endowment. (*History* 17)

The key word above is ‘hope.’ As Lindstedt claims, the purposiveness of nature and a teleological view of history are both necessary “postulates” for the possibility of the realization of the highest good and they should be seen as analogous to the postulates of practical reason (Lindstedt 143). He refers to the idea of nature working “providentially” as the first postulate (Lindstedt 143) and then goes on to say that “a second postulate...is needed so that one may hope for the achievement of a perfect morality.... [I]t may be possible for the human species as a whole, and it is this assumption which practical reason adopts” (Lindstedt 143). Lindstedt’s point here is that Kant sees nature as unfolding in a way so as to make progress toward a perfectly moral world possible, and that this achievement is indeed possible for the species as a whole.¹² Indeed it becomes

¹² What will be important however is to realize that this postulate is not a Hegelian claim that history unfolds in a way that creates moral progress, indifferent to human actions. If this were so, then any discussion of the duty to bring about the highest good on earth would be absurd. This is true for an odd reason. Earlier it seemed that a duty to promote the highest good would be illegitimate if the highest good was impossible. Here we have the opposite situation. If moral progress toward the highest good on earth occurred without human intervention, then the idea of a duty to bring it about seems redundant, as if someone were to say to us that we have a duty to make sure Saturday comes after Friday.

clear reading the *Idea for Universal History*, as well as other of Kant's works on history, that this is exactly what Kant believed.

The *Idea for Universal History* is composed of several theses. The first of these is reminiscent of Kant's discussion of the teleology of nature that takes place in the Third *Critique*. The thesis is: "*All natural capacities of a creature are destined to evolve completely to their natural end*" (*History* 18). Kant goes on to make it clear that he means both that the internal arrangement of an organism is purposive and that any given organism has a purpose within nature (*History* 18). This thesis, he notes, is necessary because it is essential to have any understanding of nature at all. Otherwise, according to Kant, "we [would] no longer have a lawful but an aimless course of nature, and blind chance takes the place of the guiding thread of reason" (*History* 18). If it is assumed that all of a creature's natural capacities are destined to evolve to their natural end, and since it is also clear that an individual human being's natural capacities cannot fully reach their natural end in one lifetime (as Kant himself claims in his second thesis), then the natural capacities of human beings must be taken as evolving as a species rather than as an individual. Kant claims that the means for this development is provided to us by nature via antagonism, or as Kant calls it, our "unsocial sociability" (*History* 21).¹³ I omit here a thorough discussion of how it is that Kant sees this "unsocial sociability" as functioning; I wish to instead focus attention on the goal to which he says it aims.

Kant notes that the goal to which nature directs and aids us is the achievement of a "universal civic society which administers law among men" (*History* 21). Kant then

¹³ For a good discussion of this issue, see Sharon Anderson-Gold's *Unnecessary Evil*.

makes the link between historical progress and nature when he says, “The highest purpose of Nature, which is the development of all the capacities which can be achieved by [hu]mankind is attainable only in society and more specifically in the society with the greatest freedom” (*History* 22). Here Kant reminds us of his earlier claim that nature aims at a final purpose, the development of the rational capacities of humankind, but he also claims here that such a development can only occur in a society that had made human freedom a realized ideal. This should remind us at once of the connection between Kant’s discussion of humanity as an end in itself and his discussion of the kingdom of ends. We must realize our true humanity, and we can only do so in a kingdom of ends.

Kant then argues that we are entitled to assume that nature indeed plays a role in helping to bring about this kingdom of ends. In a statement that should remind us of his discussion of natural teleology in the Third *Critique*, Kant asks in his seventh thesis: “Is it reasonable to assume a purposiveness in all parts of nature and to deny it to the whole?” (*History* 27). The claim here is that, just as any organism can be thought of, indeed only understood as, an organized and thus purposive being, so nature itself must be conceived of as a purposive whole (the final end of which is the development of human morality). So, Kant is arguing here that a purposiveness of nature is a reasonable supposition, and that this supposition undergirds another, that nature works to help us to bring about a condition of perfect freedom on earth.

Pauline Kleingeld picks up on this discussion of the teleological purposiveness of nature and sees in it an argument that it is God that serves the role of author of nature,

noting that, since we can regard nature as teleologically ordered (which Kant argues in the *Critique of Teleological Judgment*), we are led to the idea of God as the author of nature (Kleingeld 102). She then concludes:

Thus, if we regard nature and history as teleologically ordered, we must also regard this order as the product of a highest intelligence. This assumption makes it possible to interpret the role of God as follows: as the ground of the ‘purposive connection’ between the realm of nature and the realm of morals, God is thought to have created a world that is hospitable to moral agency.... (Kleingeld 103)

Indeed Kant himself notes that without the possibility of God as an author of nature what “would have to be surrendered in that case would be the **aim** of realizing the final end in the world (a happiness of rational beings harmoniously coinciding with conformity to the moral law, as the highest and best thing in the world) by conformity to the moral law” (*CJ* 451). This point is also made in the Second *Critique* where Kant is discussing the postulate of the existence of God. There he notes again that God is conceived of as a “cause of all nature, distinct from nature, which contains the grounds of this connection, namely of the exact correspondence of happiness with morality....Consequently, the postulate of the possibility of the highest derived good (the best world) is likewise the postulate of the reality of a *highest original good*, namely of the existence of God” (*CPR* 125). So, a picture begins to emerge that Kant sees the highest good in the world, as a moral world wherein there is happiness coextensive with morality, as indeed possible, but made possible not only through human agency but through nature guided by God. However, it is important to remember that this progress toward the moral world still requires the efforts of human beings. We can see this theme when we consider another

work on history, *The Conjectural Beginning of Human History*, as well as *Perpetual Peace*.

In *The Conjectural Beginning of Human History* Kant hypothesizes that human beings became aware early on that they stand out over and above the rest of nature, and so realized that “[human beings are] the true end of nature...” (*Beginning* 114). He writes that these conjectural beginnings of human history (which include a coming into awareness of moral imperatives) show us that history is “not a decline from good to evil, but rather a gradual development from the worse to the better; and nature itself has given the vocation to everyone to contribute as much to this progress as may be within his power” (*Beginning* 123). It is the last piece of the above quote that is important to notice. Kant claims here that, not only does nature contrive to help bring about a more moral world, but that we must contribute to this goal as well. What is interesting is that these works, purporting to look backwards, are really driven by a look forwards by the idea that everything is driven toward a world that resembles the highest good on earth or a kingdom of ends. This theme becomes clearer when we look at an additional work which is more overtly forward-looking.

In *Perpetual Peace* Kant lays out the conditions necessary for an un-ending peace between nations (which he is careful to distinguish from a simple absence of conflict). Kant’s individual prescriptions for perpetual peace are worth a look in their own right, however I will only mention two, and those only briefly, as it is in his supplements and his appendices that we best see connections between nature on one hand, and the highest good on earth on the other. Kant claims that states must meet two conditions (among

others) if they are to be able to contribute to a perpetual peace. First, he discusses a state of equality among citizens, which he describes as follows: “external (juridical) equality in a state is that relationship among the citizens in which no one can lawfully bind another without at the same time subjecting himself to the law by which he also can be bound” (*Perpetual Peace* 350n). The idea here is that all subjects must also be citizens.

Important for our purposes however is the similarity of this condition to Kant’s explanation of the kingdom of ends. Indeed the two precepts are almost identical.¹⁴ The second condition Kant lays down as giving “a favorable prospect for the desired consequence, i.e., perpetual peace” is a republican constitution (*Perpetual Peace* 351). Kant goes on to explain that this has less to do with who is in power than with how that power is administered. He talks of a division of powers between the legislative and the executive, but the motivating force behind his discussion is again the concern for protecting against tyranny and for guaranteeing that any law treats all those under the law also as legislators of the law. So, the society Kant envisions can here be seen as one that echoes the conditions necessary for a kingdom of ends. However, it is important that here (as well as in Kant’s Formula of the Kingdom of Ends itself in the *Groundwork*) the focus is not just on what the kingdom of ends will look like but on what is necessary to bring it about.

The fact that Kant talks about what it is that is necessary to bring about the kingdom of ends, or highest good on earth, is essential because it shows us that Kant does not conceive of nature as acting unaided in bringing about such a world, and indeed it is

¹⁴ See Kant’s statement that “a rational being belongs as a *member* to the kingdom of ends when he gives universal laws in it but is also himself subject to these laws” (*Gr* 434).

only by acting in conformity with the moral law (which is equivalent to acting as a legislator in the kingdom of ends) that we can bring such a world about. We must remember this, because Kant at first makes it seem like this is not true, again emphasizing the role of nature over our own role. In the First Supplement Kant writes that: “The guarantee of perpetual peace is nothing less than the great artist, nature.... In her mechanical course we see that her aim is to produce a harmony among men, against their will and indeed through their discord” (*Perpetual Peace* 361). Here Kant is echoing some of the claims we saw in his *Idea for a Universal History*. Here though, he spells some of these claims out in more detail. He makes three claims about the working of nature, intended to support his claim that nature works through discord to produce harmony:

- 1) In every region of the world, she has made it possible for men to live.
- 2) By war she has driven them even into the most inhospitable regions in order to populate them.
- 3) By these same means, she has forced them into more or less lawful relations with one another. (*Perpetual Peace* 362)

Kant goes on to flesh out these principles in a bit more depth, but the point is clear enough already. Kant’s claim, for what it’s worth, is that nature forces human beings to resolve conflicts with one another. In creating the circumstances for such conflicts to arise, nature also provides the opportunity for resolution to arise. This claim is important not for what it says about nature, but for what it says about our role in making moral

progress possible. Nature only provides the opportunity for moral progress, but only human agency can make it a reality. Kant makes this clear in the first Appendix.

In the first appendix to *Perpetual Peace* Kant takes pains to remind the reader of the role of moral duty in the promotion of such a state of peace. He begins by saying that “[i]t would obviously be absurd, after granting authority to the concept of duty, to pretend that we cannot do our duty, for in that case this concept would itself drop out of morality...” (*Perpetual Peace* 370). Here Kant is speaking to the idea that a state of perpetual peace must indeed be possible, and indeed that human agency must be able to bring it about, or else any discussion of a duty to do so would cease to make sense. Kant writes that “the practical man, to whom morality is mere theory even though he concedes that it can and should be followed, ruthlessly renounces our fond hope [that it will be followed].¹⁵ He does so because he pretends to have seen in advance that man, by his nature will never will what is required for perpetual peace” (*Perpetual Peace* 371). Kant’s argument seems to be simply that we cannot know in advance that human beings are or are not capable of willing perpetual peace. One might object that therefore Kant has equally no right to claim confidence in progress toward perpetual peace. At this point, the objection can only be countered by the claim that reason drives us to believe in such a state, such as the highest good in the world, since reason seeks an ultimate end of

¹⁵ The insertion here is Beck’s from his translation of the text, and not mine.

action, and in the absence of proof of the impossibility of such a state of affairs, we are entitled to believe in it.¹⁶

Crucially, Kant's next step is to remind us that the duty to promote perpetual peace must ultimately be rooted in a discussion of the moral law. In his words, "if this end (in this case perpetual peace) is a duty, it must be derived from the formal principle of external actions" (*Perpetual Peace* 377). The formal principle here is of course the categorical imperative, and I have already tried to demonstrate that Kant's discussion of a kingdom of ends is indeed connected to his other formulations of the categorical imperative and that his discussion of the conditions for perpetual peace echoes important aspects of his discussion of the kingdom of ends (also conceived as the highest good on earth). But Kant does not just say that our duty to promote perpetual peace is connected to the categorical imperative. In an absolutely critical passage, he goes further than this, saying that it is only by such an adherence to the dictates of the moral law that such a state of affairs is possible at all. I quote the following passage at length to underline its importance:

Then it may be said, 'Seek ye first the kingdom of pure practical reason and its righteousness, and your end (the blessing of perpetual peace) will necessarily follow.' For it is the peculiarity of morals, especially with respect to politics known a priori, that the less it makes conduct depend on the proposed end, i.e., the intended material or moral advantage, the more it agrees with it in general. This is because it is the universal will given a priori (in a nation or in the relations among different nations) which determines the law among men, and if practice consistently follows it, this will can also, by the mechanism of nature, cause the desired result and make the concept of law effective. (*Perpetual Peace* 378)

¹⁶ In the next section I argue that this is insufficient grounds for belief in the possibility of the highest good, which is why Kant must develop his view of the highest good as happiness in proportion to virtue and introduce the postulates of practical reason.

This passage is integral in reinforcing the claim that it is not by seeking such a state by whatever means we deem most efficient that we create it, but rather, by following the moral law that the attainment of conditions such as perpetual peace or the highest good become more likely.

Kant suggests that moral progress is not impossible and that it is legitimate to see humankind as progressing, with the aid of nature guided by God as an author of nature, toward the eventual possible realization of a moral world, which is a state of mutual freedom in which human beings act as autonomous agents. We have a duty thus to help bring this state of affairs about by working to advance the freedom of others and to not interfere with their free aims. This, as I argued at the beginning of the chapter, shows the connection between Kant's conceptions of the highest good or kingdom of ends and his discussion of other formulations of the categorical imperative. We have a duty of beneficence, and of cultivating our talents, which can be traced to a respect for humanity. Yet these same duties also contribute to a realm of ends, since they involve our duty to refrain from interfering in the ends of others and of helping others to realize their freely chosen ends. However, it is important to now look in more depth at the relationship between Kant's conception of the moral world, or highest good in the world, and the moral law. This is true at least in part because, as Kleingeld reminds us:

The only way to bring the sensible world into conformity with the moral world is through obeying the moral law. The idea of the moral world is *not* presented here as an attractive goal independent of the moral law, for which we subsequently have to figure out the most expedient means to attain it. Rather, the idea of the moral world is constructed on the basis of the question what a world would look like in which everyone always obeys the moral law from duty. (Kleingeld 93)

Although it is crucial for an understanding of Kant's ethics that we see it as concerned with ends, especially the end of the moral world, we must avoid the trap of going too far in the other direction. It is equally important to remember that the moral world is not *independent* of the moral law. And, as we will see in the final section of the chapter, Kleingeld's remark about the link between the moral law and the highest good will give us reason to reject the equation of the highest good as a moral world to be produced in the world of sense with the highest good as individual happiness proportioned to individual morality.

The Moral Will and the Moral Law

The claim that adherence to the dictates of the moral law is necessary to make a moral world a real possibility, and even that we have a duty to promote such a moral world, raises an interesting problem. The risk is that a reference to discussion of the highest good as an end of action either threatens to undermine the place of the moral law in Kant's ethics or that it is, as some commentators have suggested, simply a concept with no practical utility at all.¹⁷ It is important to show that Kant can avoid both of these potential pitfalls if we maintain the importance of the link between the moral law and the role human agency can play in bringing about a moral world.

¹⁷ Lewis White Beck is probably the chief proponent of this latter view. See his *Commentary on Kant's Critique of Practical Reason*, especially pp. 242-245 and particularly the comment: "The truth of the matter is that the concept of the highest good is not a practical concept at all..." (Beck 245). Thomas Auxter appears to make the same point, but later makes the case that there are two conceptions of the highest good in Kant and only one is definitively ruled out as a practical concept. Thus, when he says that "we do not use the highest good of the 'Dialectic' as the model or idea that regulates our conduct..." (Auxter 127), we must remember that he does not take this to mean that there is no conception of the highest good that can not serve as a practical guide for conduct. I take up this issue of different conceptions of the highest good in the final section of this chapter.

The first issue we must confront is whether for Kant the highest good can be thought of as a determining ground of the will without undermining the moral law.

Victoria Wike argues in *Kant on Happiness in Ethics* that indeed objective ends can be determining grounds of the will if the end works “in conjunction with or follow[s] from the moral law” (Wike 135). So, she argues, humanity as an end in itself, for example, is a determining ground of the will. This is because it works “in conjunction with, because it follows from, the moral law” (Wike 135). She then adds that the same could in fact be said to be true of the highest good (Wike 135). She goes on to make the case that indeed the highest good is inseparable from the moral law, since it “contains the moral law in that one of its elements, virtue or morality, refers to acting for the sake of duty, for the sake of the moral law” (Wike 136). This is reinforced by Kant’s frequent use of ‘worthiness to be happy’ as a substitute for virtue. In the *Groundwork* it is made clear that virtue, or worthiness to be happy, is synonymous with acting for the sake of the moral law. So Wike shows us that there is nothing problematic with the view that the highest good is the determining ground of the will, as long as we realize that it is so “together with the moral law, because it contains the moral law and is an object dictated by the moral law...” (Wike 136). The highest good, or any objective end, the promotion of which is commanded by the moral law does not undermine the moral law as long as we realize that such ends must follow from and be prescribed by the law. This may seem like a small point, but it is important to remember that we cannot say that duties, such as one to promote the highest good or the kingdom of ends, or any other end such as the happiness of others or the cultivation of our talents, are duties because of the goodness of

the end pursued, independent of the commands of the moral law. However, if the highest good does not undermine the commands of the moral law, there is still the risk that it is an unnecessary and practically useless concept, for the very reason that it follows from the moral law.

Lewis White Beck perhaps puts the point most poignantly when he asserts that, when we are confronted with the question of what we are to do to promote the highest good we “simply act out of respect for the law, which I already knew” (Beck 244-5). In some sense, Beck must be right when he makes this point. In fact it is crucial to my view that acting to promote the highest good is not separated from acting as legislators in a kingdom of ends, or from treating humanity as an end in itself, or from acting so that our maxims can conform to universal law. However, what I wish to argue here is that this fact does not in fact lead to Beck’s conclusion, that “the concept of the highest good is not a practical concept at all” (Beck 245). The concept of the highest good may not give us any new duties, but this does not necessarily mean that it cannot be *practically* useful as a concept.

Kant’s explication of the categorical imperative itself gives us good reason to believe that there is frequently something to be gained from recasting a concept in a new way, even if the concept remains logically the same. Kant makes this point when he discusses the various formulations of the categorical imperative. He notes that there is in fact an important difference between the various formulations of the categorical imperative, but that this difference “is indeed subjectively, rather than objectively practical, intended namely to bring an idea of reason closer to intuition.” (*Gr* 436) He

then goes on to emphasize this point, writing that if “one wants to provide *access* for the moral law, it is very useful to bring one and the same action under the three concepts mentioned above and thereby, as a far as possible, bring it closer to intuition” (*Gr* 437). Clearly then, there is a great deal to be gained from recasting the moral law in different ways, so that we are better able to apprehend what is indeed entailed by the categorical imperative.

This point impacts the relationship between the highest good and the categorical imperative in an important way. Just as the various formulations of the categorical imperative are given expressly for the purpose of helping us access the moral law, and thus our duties, the highest good can be seen as serving a similar purpose. The various formulations of the categorical imperative offer no duties that were not entailed by the version Kant first puts forth. Nonetheless, the various formulations help make it possible for us to *see* how to think about the moral law in ways that may have been closed to us if we only had access to one formulation. Similarly, the highest good can help us to see the full breadth of our duties in a way that might be obscured by relying exclusively on a conceptual analysis of the categorical imperative.

Several authors have noted the fact that indeed a focus on the promotion of a moral world or highest good in the world changes the way we might think about our duties. For example, Pauline Kleingeld notes that:

With the determination of the highest good as the final end or object of all moral agency, the horizon of a moral agent shifts. The will extends...itself *from* the narrow perspective of the immediate duty to act on maxims that one can at the same time will to become universal law, *to* a broader perspective, from which all

moral actions are seen in the light of the one final end to which they contribute” (Kleingeld 95).

The idea of a shifting of one’s moral horizon is helpful, because it serves to show us that, although we may still test our maxims by reference to universality, we can also use a different lens that allows us to see our moral actions as contributing to a final end and thus as *leading somewhere*.

Victoria Wike takes a similar approach to the one Kleingeld mentions and which I am advocating when she notes that “the duty to seek the highest good certainly commands something different but probably does not command something more [than is commanded by the moral law]” (Wike 142). This is because, as Kleingeld suggests in the passage above, the moral world is a world in which there is cooperation among moral agents; it is, in Wike’s words, “a social object or end” (Wike 143). Thus any understanding of the highest good helps us to realize that adherence to the moral law does two important things: it helps make a moral world possible, and it reminds us that we have a moral obligation to bring it about.¹⁸ These are both features of our moral lives that would be much harder to see if not for the concept of a highest good in the world.

Authors such as Jennifer Moore, Sharon Anderson-Gold and Pauline Kleingeld take up the issue of what sorts of duties are revealed to us when we make use of the concept of the highest good in the world, though I mention them here only to emphasize the point that these descriptions of the duties entailed in a duty to promote the moral world would perhaps not occur to us if we thought of our moral responsibilities solely in

¹⁸ Silber’s discussion of the highest good as a regulative obligation in “Kant’s Conception of the Highest Good as Immanent and Transcendent” is useful here, since Silber argues that it is the goal of attaining the highest good that provides us with something to strive toward.

terms of the categorical imperative. Moore notes, for example, that an ethical community would consist of the features of unity, harmony, a commonly held valuing of the community itself, value for fellow members of the community that includes affective or moral bonds, reciprocity, and a source of the identity of the individual members with the community, and she goes on to argue that Kant's notion of an ethical community includes all of these features (Moore 59-60). Anderson-Gold speaks specifically about our duties in what she calls the ethical commonwealth, writing that duties "of virtue go beyond viewing others as limits to our claims and directs us to treat others as 'ends' supporting and promoting their legitimate objectives....It entails minimally the abandonment of aggressive and competitive attitudes toward others, and maximally the adoption of cooperative and supportive networks" (Anderson-Gold 30). Finally, Kleingeld takes care to distinguish Kant's conception of a moral world from any sort of political community, an issue which Kant takes up in some depth in the *Religion* as well as in *Perpetual Peace*.¹⁹

The idea of the moral world, like that of the kingdom of ends and humanity as an end in itself, helps provide a lens which can offer accessibility to the abstract command to follow the moral law or to will only those maxims that can serve as universal law. However, such notions, especially that of the moral world, offer us something else that is

¹⁹ My brief use of examples from Moore, Anderson-Gold, and Kleingeld does not do justice to the depth of their discussions of what an examination of the moral world can tell us about our duties. I recommend Moore's article "Kant's Ethical Community", Anderson-Gold's article "Kant's Ethical Commonwealth: The Highest Good as Social Goal," and Kleingeld's "What do the Virtuous Hope For? Re-reading Kant's Doctrine of the Highest Good" if the reader is interested in these issues. My point here is more basic, that an appreciation of Kant's discussion of the moral world helps us to see that adherence to the moral law can be seen as contributing to such a world.

equally important. By seeing that adherence to the moral law is at the same time a duty to, for example, promote the moral world, we are able to provide a further answer to the question: Why must I refrain from lying, even to a murderer at the door? The fact that the moral world is connected to the moral law allows us to hope, if not to assume, that adhering to the dictates of the moral law is connected to making a moral world possible. In the final chapter I defend this as a possible answer to those who think Kant must lower his moral standards in the face of extreme circumstances. However, before concluding our discussion of the highest good as a moral world, something should be said about its relationship to Kant's discussion of a related notion of the highest good that is present in Kant's thought, namely as happiness proportioned to morality, possible only in the afterlife. This discussion is necessary because of two possible problems presented by what Andrews Reath calls the 'theological conception' of the highest good²⁰: first, many critics have dismissed Kant's notion of the highest good because of problems with this theological notion, and second, unlike the idea of a moral world, the idea of a duty to apportion individual happiness to individual virtue in this world is not in keeping with Kant's understanding of our moral obligations, so we should be careful not to see the connection between the moral law and the highest good as being tied primarily to this theological conception.²¹

²⁰ See Andrews Reath, "Two Conceptions of the Highest Good in Kant"

²¹ Dusing attributes much of the confusion over the various roles the concept of the highest good plays in Kant's thinking to a failure to appreciate the fact that the questions with which Kant is dealing, regarding the highest good, change throughout his work. While this seems obvious, many commentators have tried to find a unified view of the highest good and Dusing correctly suggests that this is misguided. Dusing points out that the discussion of the highest good as personal happiness proportioned to personal virtue can be

Problems with the Proportionality of Happiness and Virtue

As discussed above, Kant often associates happiness conditioned by morality with the moral world, which he also calls the highest good in the world. However this association of happiness and morality is given various articulations by Kant in his writings, and, while I have so far only focused on descriptions of the highest good that emphasize the role of the highest good as the social ideal of a moral world, many of Kant's other expressions of his notion of the highest good are more ambiguous and seem to suggest that Kant employed more than one notion of the highest good, or at least that the highest good serves multiple roles within his thinking. One particularly interesting passage occurs in the Dialectic of the *Critique of Practical Reason*. There Kant writes that "virtue and happiness together constitute possession of the highest good in a person, and happiness distributed in exact proportion to morality (as the worth of a person and his worthiness to be happy) constitutes the *highest good* of a possible world..." (CPR 111).²² The passage thus seems to suggest two related notions of the highest good, or that Kant is placing it simultaneously in two different roles.²³ On the one hand, we see Kant refer to

traced to the fact that Kant is taking up a question that ancient philosophers had discussed: whether happiness was the condition for virtue or vice versa. He discusses this issue at length in part one of his article, "Das Problem des Höchsten Gutes in Kants Praktischer Philosophie." But, Dusing goes on to suggest that Kant comes to see the highest good much more in terms of a moral world. For example, he points out that "Kant bestimmt in der Aufnahme der christlichen Tradition das höchste Gut als intelligible Welt, als Reich Gottes oder als 'moralische Welt'. Im Unterschied zu den Idealen der Alten, in denen die Vollendung des einzelnen Menschen vorgestellt wird, ist damit das höchste Gut ein ethischer Weltbegriff. Es wird als ein Ganzes vernünftiger Wesen gedacht, in dem Sittlichkeit und Glückseligkeit harmonisch zusammenstimmen" (Dusing 17). What Dusing does not make clear here is whether the moral world is one possible in the world of sense.

²² This passage lends support to the view that Kant did not abandon one conception of the highest good for another at some point within his writing, but rather that he saw these roles or conceptions of the highest good as existing at the same time.

“the highest good of a possible world.” This reminds us of other passages in which the highest good is conceived of solely in these terms²⁴. Yet, in the very same passage, we see mention made of the “highest good in a person” and we also see the addition of the requirement of “proportionality” between happiness and virtue, which is largely absent from many of Kant’s other passages about the highest good in the world.

It seems clear that Kant does in fact see the highest good as happiness proportioned to morality within a person as something that is possible only in the after life. This is made clear when Kant postulates the immortality of the soul as a necessary condition for the possible attainment of such a highest good. He does so because he notes that “the *complete* conformity of dispositions with the moral law is the supreme condition of the highest good” (*CPR* 122). But of course such a complete conformity with virtue seems like it cannot be expected (though this does not mean it is impossible) within one individual’s lifetime, so we require the possibility of “an *endless progress* toward that complete conformity...” (*CPR* 122), and thus “the highest good is practically possible only on the presupposition of the immortality of the soul...” (*CPR*122). It is interesting to note here that this ‘endless progress’ only requires the postulate of the immortality of the soul if it pertains to the highest good within an individual. As we have already seen, such progress can take place through history if the progress is toward the

²³ Victoria Wike makes this point on p. 128 of *Kant on Happiness in Ethics*.

²⁴ For example, as we saw earlier, in the Third *Critique* Kant describes the highest good in the world as “happiness – under the objective condition of the concordance of humans with the law of morality, as the worthiness to be happy” (*CJ* 450). Additionally, in the First Critique Kant describes the highest good as a world in which “freedom, partly inspired by and partly restricted by moral laws, would itself be the cause of general happiness, since rational beings, under the guidance of such principles, would themselves be the authors both of their own enduring well-being and of that of others” (A 810/B 838).

social goal of the highest good as a moral world. It seems reasonable then to assume that, to the extent that Kant is dealing with different notions of the highest good, it is the difference between the highest good within a person and the highest good within the world that is being worked out.²⁵ This seems true despite the fact that Kant also seems to largely associate the ‘theological’ conception of the highest good with proportionality, while he uses this phrasing less often to describe the highest good as a moral world.

It is tempting to suggest that the addition of the requirement of proportionality is what distinguishes the two notions of the highest good, and this at first seems likely. For example, Reath makes the point that the union of happiness and morality or general happiness and general virtue is not the same as proportioning happiness and virtue, since the latter implies a necessary connection, or causal relationship, between the two and the former does not (Reath 604-5). This point seems to be born out by the fact that Kant’s only extended discussion of any causal relationship between happiness and virtue takes place within the Dialectic of the *Critique of Practical Reason* and it is there that he is discussing the question of proportionality. However, this issue is not without controversy. The passage that makes such a view troublesome occurs when Kant alludes, in the First *Critique*, to “the moral world...in which happiness as bound up with and proportioned to morality....” (A 109/ B 837). Here Kant makes explicit reference to a moral world, yet refers to proportionality. This leaves open the possibility that Kant sees

²⁵ I would suggest this is true even though Kant’s language seems less clear. For example, in the passage in which Kant argues for the postulate of the immortality of the soul he uses the phrase “the highest good in the world” (*CPR* 122). However, it is clear from the rest of the passage that he is clearly referring to the highest good within an individual. It is not clear from Kant’s own writing whether he saw a clear divide between the issue of the highest good within a person and the highest good in the world.

proportionality of happiness to virtue as a possibility even within the world of sense.²⁶

Since the question is open, I would suggest that we are better off thinking of the difference between the two notions in terms of the fact that the ‘theological notion’ is one in which happiness is proportioned to virtue *within the individual person*.

This distinction is important because, even if we grant that Kant thought of the relationship between happiness and virtue in a moral world as one that involved proportionality, we would not be committed to the view that this means the moral world is one in which each person was happy to the degree that they were virtuous. As we will see in a moment, this is an important point because, as Kleingeld points out, “Kant’s teleological account of history may indicate how a moral world may be approached, but it does not account for a possible *proportional* allotment of happiness according to individual worthiness to be happy” (Kelingeld 105). So, if Kant *was* espousing a moral world that is possible in the world of sense, he either could not have intended that it contain the requirement of proportioning individual happiness to individual morality or, if he did intend this, he failed to establish the possibility of such a world.

We seem then to have two views of, or roles for, the highest good: one is social and is conceived of as possible in the world of sense, and one is personal and is conceived of primarily as possible only in the afterlife. At a minimum then, this helps us to see that in fact Kant accepted the view of the highest good as a moral world in addition to one of individual happiness in proportion with individual morality, and the earlier part

²⁶ Reath considers this point in “Two Conceptions of the Highest Good in Kant” but doesn’t pursue it. He notes: “Proportionality could be the basis of a secular ideal, and Kant may occasionally adopt such a view. To simplify, I will ignore this possibility and will treat proportionality as the description associated with the theological interpretation” (Reath 603).

of the chapter set out to make this case. But more than this, it is important to see that the conception of the highest good as a moral world is not necessarily subject to some of the problems associated with the highest good as happiness proportional to morality within the individual. Andrews Reath puts the point this way: “The critics have thought that by pointing to ...[the] flaws [in the conception of the highest good as happiness proportionate to virtue], one could dismiss the entire doctrine” (Reath 594). I have tried to argue that it is important that the highest good not be dismissed as an important part of Kant’s ethical thought, primarily because understanding a connection between the highest good understood as moral world and the moral law helps us to see that our adherence to the moral law can be seen by us as helping contribute to progress towards a more moral world. My goal here is then not to defend the highest good as happiness proportioned to virtue from criticism, but rather to make it clear that any criticisms of such a notion of the highest good need not undermine the highest good understood in the sense I have been discussing it, as a moral world. I would in fact suggest, in keeping with critics such as Thomas Auxter and in opposition to John Silber, that the highest good conceived of happiness in proportion to virtue should not be seen as a guide for our conduct.

The title of Thomas Auxter’s article “The Unimportance of Kant’s Highest Good” suggests a far more thorough rejection of the notion of the highest good than he actually offers. Auxter’s argument is in fact only that the notion of the highest good as happiness proportioned to virtue ought to be rejected as a guide for our moral conduct. Auxter in fact acknowledges the existence of two notions of the highest good within Kant, differentiating the moral world, which he refers to as the highest good as ectypal world,

from the highest good as happiness in proportion to virtue, which he refers to as an “otherworldly condition for which a virtuous person might hope” (Auxter 127).²⁷ Auxter notes that there is a “prima facie case for identifying the ectypal world (the ideal world of sense) with the highest good of the ‘Dialectic’ and thus for holding that this highest good *is* suitable as a standard for guiding conduct” (Auxter 125), but he goes on to reject this notion. The basis for his rejection is quite clear. Any duty to bring about the highest good must assume that it can indeed be brought about by human agency in this world of sense, and thus a description of the highest good “cannot refer to powers human beings would have if they were not subject to the constraints imposed by living in an empirically conditioned world....It might be that one who had divine powers could create a world in which happiness is proportionate to virtue. But this is not a choice open to human beings” (Auxter 125).²⁸ Auxter is clearly correct here. In order to bring about the moral world, we need only try to promote conditions such that agents can maximally realize their autonomy and their freely chosen ends. Such a world requires virtue from us and gives us reason to believe general happiness will result. Even here, we need to assume the existence of a moral author to help guarantee the actual production of happiness, nature must be ‘hospitable’ to moral action, though far more is asked in any attempt to proportion happiness to virtue within the individual. For one, we cannot know any

²⁷ Auxter leaves open the possibility that the highest good as ectypal world is in fact a valid practical concept and even links it to the categorical imperative, saying “When we universalize our maxims, we are testing them to see whether they are fit to become part of the content of this model [of the ectypal world]” (Auxter 126).

²⁸ Lewis White Beck makes this same point in his *Commentary on Kant’s Critique of Practical Reason* when he writes that the task of apportioning happiness to virtue is the task of “a moral governor of the universe” (Beck 245).

agent's worthiness to be happy (virtue), and so any duty to proportion happiness based on this worthiness is beyond the possibility of human agents.

The issue of the inability to proportion happiness to virtue leads to a practical reason for thinking that the highest good as happiness proportional to virtue is not to be taken as a practical principle in the way that the highest good as moral world is. We can see this when we note exactly what it is that Auxter and Beck are responding to when they take issue with the highest good, as happiness in proportion to virtue. Both authors' treatments of the highest good come in large response to a series of articles by John Silber on the highest good. In "The Importance of the Highest Good in Kant's Ethics," for example, Silber makes brief mention of the fact that we do, in many situations attempt to promote happiness in proportion to virtue. He notes, "in rearing children, serving on juries, and grading papers one tries to do and actually can do something 'about apportioning happiness in accordance with desert'" (Importance 183). Silber's approach seems promising at first, and it seems that indeed his approach can be used in other examples as well. It seems that we can conceive of a way of conducting business that rewards employees for ethical conduct rather than, or in addition to, rewarding them for their ability to maximize profits. We can also imagine a political structure that rewards ethical behavior in the private sector. To an extent, we could even see the fact that charitable contributions result in tax deductions as a step in this direction. In international politics too, we often see governments and international bodies such as the UN imposing sanctions on governments that violate human rights. However, when we

examine this issue further, we see that it raises several concerns about the concept of merit.

Whenever we try to use the appearance of merit for the condition for allocating a reward we run several risks. First, we run the risk of creating the reward as the incentive for action. However, there is nothing in any of the above examples that necessarily implies that the moral conduct that is being rewarded was undertaken for the sake of the reward. People give to charities for any number of reasons, and all of these possible motives are hidden from the outside observer. It is clearly possible that people who give to charities act on the maxim of beneficence and not from any desire for a reward. Yet, this leads to a second problem, that of presuming that we have insight into an individual's actual motivation for action. If we were responsible for apportioning individual happiness to individual virtue, the duty would be to reward virtuous dispositions, not just actions that *seem* to indicate the presence of such virtue.²⁹

It is clear that we cannot possibly have a duty to apportion happiness to virtue in a given agent if we can never have a sense of the agent's actual worthiness to be happy. Such a duty would push us back in the mire of having a duty to do the impossible. It would be impossible to apportion happiness to virtue because it is impossible to ever truly assess virtue. But, even if we could assess virtue, allocating happiness on this basis

²⁹ Stephen Engstrom makes note of this problem in his article, "The Concept of the Highest Good in Kant's Moral Theory", especially in section III. His solution is one I have already alluded to: we 'apportion' happiness to virtue by promoting each of these independently of an agent's actual worthiness to be happy at any given time. Engstrom's argument is that we cannot know their overall worthiness to be happy, since we cannot know their motives or their futures (and often even their past). Yet, since virtue is the necessary end of a rational being, I can still 'proportion' happiness to virtue by acting so as to promote an agent's happiness, since they are already responsible for promoting their own virtue.

seems to lead to absurd conclusions. Consider, for example, any occasion when we decided to contribute to a person's misery simply because we thought them to be lacking in virtue. This seems clearly contrary to the spirit of Kant's ethics. Equally contrary to the spirit of Kant's thinking is the idea of trying to make someone happy simply because we think they seem like a moral person. Doing so would place us in a position of being motivated by a person's worthiness to be happy rather than by duty. These considerations thus give us good reason to reject the notion of the highest good as happiness in proportion to virtue within an individual as one that is to serve as a practical principle guiding our action and providing us with duties.

We thus have two reasons to reject the notion of individual happiness proportioned to individual virtue as a practical principle we use to govern our actions and understand our duties. In the first place such a view would undercut any duty to bring about the highest good, since a highest good conceived in this way would be beyond human powers, not just of any individual person, but of the species as a whole.³⁰ Second, there are problematic practical implications regarding what would follow from any alleged duty to promote individual happiness proportionate to individual morality. Most troubling is the fact that such a duty seems to undercut the purity of the moral law by making our moral motivation bound up with another person's moral merit rather than being based on a respect for the unconditional commands of the moral law. However, the

³⁰ Although Silber, in "Kant's Conception of the Highest Good as Immanent and Transcendent," suggests that we can have the duty to promote the highest good while using the ideal of the actual attainment of the highest good as a regulative idea that helps promote continual striving, this argument does not seem to work if the attainment of the highest good is not ever possible within the world of sense.

highest good conceived of as a moral world, in which freedom allows for people to set and realize ends, does not suffer from these same problems.

Conclusion

I have attempted to show that the promotion of a moral world holds an important place in Kant's ethical thought and further. I have tried to show both that Kant was deeply concerned throughout his writings about the idea of progress towards this moral world and that such a world can indeed be conceived of as possible in a world of sense. But most importantly, I have tried to point out the fact that, to the extent that human agency helps create this moral world, it is through adherence to the moral law that we do so.

There are two dangers that we must avoid in thinking about the role of the moral world in Kant's thought. The first is that we dismiss it as a minor or irrelevant aspect of Kant's thought. The second is that we over emphasize its importance to the degree that we come to see the promotion of a moral world as an end to be achieved but we lose sight of Kant's guidance as to how it is to be brought about. Those who emphasize the *Groundwork* and the categorical imperative as the keystones of Kant's ethics run the risk of falling into the first trap, and this chapter and the preceding chapter are both aimed at preventing this. On the other side of things, however are those who see in Kant's descriptions of a moral world an ideal state of affairs that ought to be brought about, but who would suggest that we cannot bring such a world about while holding ourselves bound to the strict commands that seem to be generated by the categorical imperative. It is to this latter set of thinkers that we will turn our attentions in the final chapter.

CHAPTER FIVE

THE PLAUSIBILITY OF KANT'S APPROACH

This dissertation ends where it began, by looking at Christine Korsgaard's treatment of Kant's views regarding whether or not it is permissible to lie to a murderer at the door. In chapter one of this dissertation I examined Korsgaard's claim that, while the categorical imperative forbids lying to the murderer at the door under the Formula of Humanity, the lie could be justified under the Formula of Universal Law. I argued there that this was not the case, arguing Kant's understanding of what is entailed in willing or thinking a maxim that would contradict itself when applied universally was not as Korsgaard depicted it to be, and thus that the various formulations of the categorical imperative yield the same results in cases such as the murderer at the door. But there is a second part of Korsgaard's thinking about the case of the murderer at the door, and in general what she refers to as dealing with evil, that motivates the current chapter. Korsgaard's argument is really motivated by the view that we *ought* to separate the formulations of the categorical imperative since the standard of the Formula of Humanity and the Formula of the Kingdom of Ends are unrealistically stringent when it comes to dealing with evil. She writes that:

The Formula of Humanity and the Kingdom of Ends will provide the ideal which governs our daily conduct. When dealing with evil circumstances, we can depart from this ideal. In such cases we can see that the Formula of Humanity is inapplicable because it is not designed for use when dealing with evil. But it can still guide our conduct. It defines the goal toward which we are working, and if we can generate priority rules we will know which features of it are most important. (Korsgaard 151)

Korsgaard thus represents a sort of ‘middle ground’ between Kant’s ethics and non-Kantian approaches to ethics, such as that taken by utilitarianism or feminist ethics.

While these other approaches might reject Kant’s own line of thinking from the outset, Korsgaard represents the view that we ought normally make use of Kant’s approach to morality, but that we need to make room within that approach for dealing with situations wherein adherence to the moral law seems like it asks too much of us.

Thus there are two main goals of this chapter: first, I wish to offer some possible lines of defense against the general charge that his rigorism is implausible, especially when it comes to situations in which we feel a tension between a desire to prevent tragic outcomes and an adherence to the moral law. I will explore two different types of criticism of Kant’s approach here. The first is concerned with what we can call cases of ‘special circumstance,’ where the basic claim is that Kant’s rigorism simply cannot deal appropriately with a certain kind of scenario; in this type of scenario we can see Kant’s approach to morality as facing a challenge from consequentialism. The second type of challenge to Kant deals with the view that it is the specialness of individual relationships that causes a conflict with Kantian morality. Here the challenge to Kant can be seen as coming from a feminist perspective on ethics. Then, after dealing with these challenges

to Kant's approach to morality, I turn my attention to Korsgaard's specific approach. I wish to argue that her approach, which employs a 'double-level theory', amounts to an adherence to a Kantian morality except when the 'going gets tough' and is untenable.

It may be helpful then to begin by contrasting Kant's approach with a few other approaches, focusing simply on the issue of what we are to do when our desire to prevent a tragic outcome comes up against a desire to adhere to rules, laws or principles. My goal here is not to offer a general critique of non Kantian approaches to morality, nor is it to engage in a full debate with those who would reject a Kantian approach out of hand. My aim here is more modest in that I simply wish to challenge Korsgaard's specific reasons for uneasiness with Kant's approach. Korsgaard notes that the "trouble is that in cases such as that of the murderer at the door it seems grotesque simply to say that I have done my part by telling the truth and the bad results are not my responsibility" (Korsgaard 150). I would argue that the characterization of such an attitude as "grotesque" dismisses it prematurely and presumes that Kant's reasoning is simply indefensible. I therefore wish to begin the discussion by offering some considerations that might prevent us from dismissing his approach, even in situations such as the murderer at the door, too easily.

It will be helpful then to begin by examining a few scenarios in which the tensions between a desire to prevent a tragic outcome and the desire to adhere to the moral law (or laws or rules more generally) is pronounced. I would like to offer three such scenarios, each intended to bring out a different issue that might help us better appreciate the plausibility of a Kantian approach. The first issue I will examine is the

question of whether we should torture to prevent a ‘ticking bomb’ from going off. Next, I examine the fictional case of Melville’s *Billy Budd*, where we see a scenario in which adherence to the rules seems to lead to a particularly tragic event. Finally, I consider the question of whether physicians should lie on an insurance form to help ensure that a patient receives coverage and thus treatment. While all of these cases pit a respect for lawfulness against a concern for consequences, each is intended to bring out a different sort of concern about Kant’s approach to ethics, and thus each case will offer us a slightly different line of defense against the critics of Kant, who take the view that such a moral rigorism simply cannot be defended.

Torture and the Contingency of Empirical Conditions vs. the Certainty of the Moral Law

In his article “Liberalism, Torture, and the Ticking Bomb,” David Luban walks through the five general reasons people or societies have engaged in torture: torture as the indulgence of a victor over a defeated enemy, torture performed in order to terrorize a people into submission, torture used as a form of criminal punishment, torture as a means of extracting confessions, and, finally, torture used as a method of intelligence gathering¹. He goes on to note that a society with respect for rights inevitably finds torture abhorrent especially because of the way it degrades and terrorizes the individual. Yet Luban notes with alarm that our culture, living in a post 9/11 world, is far more at ease with torture for the purposes of intelligence gathering than it is with the other possible motives for torture.

¹ See Luban, pages 1432 and following.

Luban notes that the reason torture for intelligence gathering purposes has not been met with the same kind of disdain as have the other motives for torture is because the torture is used “to prevent a catastrophe...[and] that those who inflict torture are motivated solely by the looming catastrophe, with no tincture of cruelty; that torture in such circumstances is little more than self defense...” (Luban 1439). He goes on to point out that the debate over torture as a part of the ‘war on terror’ has focused almost solely on the scenario of the ‘ticking bomb,’ which he describes the following way:

Suppose the bomb is planted somewhere in the crowded heart of an American city, and you have custody of the man who planted it. He won’t talk. Surely...we shouldn’t be too squeamish to torture the information out of him and save hundreds of lives. Consequences count, and abstract moral principles must yield to the calculus of consequences.² (Luban 1440)

We can see at once that we have here an example of a modern version of the murderer at the door. Arguments in favor of using torture for intelligence gathering trade on the idea that unbending allegiance to moral principles is indefensible when such an allegiance may prevent us from saving hundreds, maybe even thousands of lives. However, the question we must ask, following Luban, is whether this is an accurate portrayal of the case at hand. Luban does well to portray the argument for torture for what it is: an argument based on the belief that consequences of actions are within our control, when the reality is that this is clearly not true. This is perhaps the first principal line of defense we can offer for Kant against those who criticize his rigorism in favor of a concern for

² To prevent misunderstanding of Luban’s argument, I should point out that he is not endorsing this view. The passage is intended to simply portray the basic argument that those who favor torture in such circumstances employ.

bringing about or avoiding certain consequences: these consequences are far less within our control than we might imagine.

Luban notes that the ticking bomb scenario is in fact highly unrealistic (Luban 1442), and a little reflection suffices to show that his assessment is obviously correct. It would be extremely rare that we could be sure the following conditions were met before engaging in torture: 1. The person we have in custody in fact has the information we need. 2. The person will give us the information we need if we torture him. 3. The person will not give us the information under any other circumstances. 4. The information could generally not be gathered by other means (such as following other leads). 5. The information will be obtained in time to save the lives we intend to save. In fact, the point should be put even more strongly: we *cannot* know the above five things for sure. As Luban points out, this is true even if torture in hindsight proves to have helped in preventing an attack³. Luban thus arrives at the inevitable conclusion that “in a world of uncertainty and imperfect knowledge, the ticking-bomb scenario should not form the point of reference” (Luban 1444). In cases such as the ticking-bomb scenario we act as if we are able to have knowledge we simply cannot have, and we act as if we have powers we simply do not. For, even if we could meet all the above conditions before deciding to engage in an act so contrary to the moral law as torture is, there is still the chance that we would get the bomb not a minute before it goes off, but a minute after. This recognition

³ Luban recounts a story in which Philippine police torture a Pakistani bomb maker in order to thwart an attempt to bomb airliners and to assassinate the Pope. The plot was indeed uncovered in time, and the information gathered from the torture was important in uncovering the plot, but, as Luban points out, the case is misleading because the torture uncovered the plot; the plot was not known at the time of the decision to torture. (Luban 1441-1442)

of the limits of the control we have over consequences is one of the important dimensions of Kant's approach to such problems.

In the introduction to this dissertation I mentioned the case of a police officer who engaged in torture to save a young girl trapped in an abandoned car. The case seems at first to be a powerful one in arguing for the claim that the ends justify the means, and as such it has a rhetorical force similar to the ticking bomb scenario. However, the problem in the case is similar to the issues Luban points out in the ticking bomb scenario. There are factors beyond our control, and it is absurd to argue that the morality of our actions depends on these factors. So, one must be able to show that the torture in either case is justified or unjustified, regardless of whether or not the bomb goes off or whether or not the girl is saved. Those who argue in favor of torture in extreme situations always presuppose that the torture in fact results in 'saving the day'. But what if it did not? Would it still be justified? If the answer is 'no,' then the plausibility of the argument to torture collapses, because then the morality of an agent's actions, the decision to torture, rests on factors such as how much air is in a car or whether traffic is such that I can get to a bomb in time, which are both conditions about which the torturer has no control and for which he or she deserves no moral praise or blame. If, on the other hand, the answer is 'yes,' then those who would argue in favor of torture are placed again in a precarious position, of having to argue that clear violations of morality would be justified in the mere attempt to prevent serious harms from occurring. Given this, it seems more plausible to argue that, rather than torture being justified regardless of results, it is

unjustified regardless of results. And part of what helps us to see the power of this claim is the acknowledgement of just how out of our control results truly are.

The above consideration of the issue of torture is intended to bring out the important practical ethical concerns that emerge when we realize that consequences and other empirical concerns are largely out of our control. I take it as well established that Kant has metaphysical grounds for rejecting a morality based on empirical conditions; he tries to ground his notion of duty in reason and lawfulness because only then can duty really carry with it a notion of obligation and thus necessity. But, it is important to see that Kant seems to be also aware that there are practical problems with a morality based on a calculation of consequences. We can see such an awareness of this issue, for example, in the *Supposed Right to Lie*.

In the *Supposed Right to Lie*, Kant, in addition to his other arguments against lying to the murderer at the door, also addresses the concern for the possible results that would follow from truthfulness. There he notes: “It is still possible that, after you have honestly answered ‘yes’ to the murderer’s question as to whether his enemy is at home, the latter has nevertheless gone out unnoticed, so that he would not meet the murderer and the deed would not be done...” (*Supposed Right* 427).⁴ Kant is here pointing to the fact that in reality, all we can control is our own choices, and we portray the picture falsely when we depict it as one in which we choose between a moral principle and a certain tragic outcome we wish to avoid.

⁴ It should be noted that Kant goes on to argue that the person who told the lie would be liable in a court of law for the consequences of the act. I omit the end of the passage here simply because my purpose is not to give an account of Kant’s discussion of ‘juridical right’ in such cases, but only to illustrate his awareness of the fact that, whether we tell the truth or not, there are other variables that are beyond our control.

Kant is able to show us that we put the equation backwards when we try to derive what is right from a concern for consequences. Consequences are beyond our control and cannot offer us a stable guide to what is right. The moral law, by contrast, gives us a consistent guide to deciding whether or not an action is morally acceptable. But, equally important is the idea that acting from the moral motive, or the good will, helps us to ensure that the morally correct actions are indeed performed. Ironically then, it is not concern for consequences that helps us determine what is right, but rather it is the motivation of acting out of respect for what is right that helps bring about moral actions. We can see this idea in Kant's discussions of the instability of inclinations, as contrasted with the stability of acting from duty. Two such passages occur in the *Groundwork* during Kant's discussion of the examples of duties stemming from the Formula of Universal Law of the categorical imperative.

Kant, in speaking about the duty to preserve our life, notes that people will, for the most part preserve their lives regardless of any duty to do so. He writes:

[T]o preserve one's life is a duty, and besides which everyone has an immediate inclination to do so....They look after their lives *in conformity with duty* and not *from duty*. On the other hand, if adversity and hopeless grief have quite taken away the thirst for life, if an unfortunate man, strong of soul and more indignant about his fate than despondent or dejected, wishes for death and yet preserves his life without loving it, not from inclination or fear but from duty, then his maxim has moral content. (*Gr* 398)

What is of interest to me here is not the point Kant is trying to make, that actions must be done from duty to have moral worth, but rather the practical benefit of having acted from duty. If we act based on inclinations, then we are at risk of taking our own lives whenever we feel that life has gotten the better of us. However, if we act on the basis of

our duty, we will preserve our lives in the face of this difficulty. So, our respect for duty provides us with a guide that is more stable than the whims of our inclinations. This point is more important when we see how it affects the way that we treat one another, in the duty of beneficence.

To help further make his point about the difference between acting from duty and merely acting according to duty, Kant then makes use of an example of a philanthropist who no longer feels any desire to be beneficent. He writes:

Suppose then, that the mind of this philanthropist were overclouded by his own grief, which extinguished all sympathy with the fate of others...; and suppose now, when no longer incited to it by any inclination, [he] nevertheless tears himself out of this deadly insensibility and does the action without any inclination, simply from duty; then the action first has its genuine moral worth. (*Gr* 398)

Here again, Kant is not himself arguing that the good will is important *because* it helps guarantee that we act morally. His point is simply to show us the difference between actions that reflect the good will and are thus done from duty and those that are done merely according to duty. Nonetheless, implicit in his example is the recognition that the instability of our inclinations tempts us away from our duties, and the moral disposition is the best way to guarantee moral behavior. Kant then passes on to a consideration of a case in which “nature had put little sympathy in the heart of this or that man...; he is by temperament cold and indifferent to the sufferings of others” (*Gr* 399). Here again, the fact that such a man nevertheless acts in a beneficent way, simply out of respect for the law is presented by Kant simply as a way to demonstrate the esteem with which we should hold such a motive. Yet again, the example invites us to ask a question: If not for

the motive of acting from duty, what would become of a person who is “by temperament cold and indifferent to the suffering of others”? The answer is clear. Without a stable motive for doing one’s duty, we run the risk that the duty itself will not be performed. So, while Kant is not here talking about how it is that we decide what our duty in fact is, he does give us reason to conclude that one benefit of his approach is that, assuming we know what our duty is, the moral disposition provides a constant reason for performing it.

This awareness of the practical consequences of a lack of a stable source of moral motivation appears elsewhere in the *Groundwork* as well. One could read the example of the shopkeeper who does not cheat his or her customers in a similar fashion. If the shopkeeper does not cheat his or her customers because of a natural feeling of affection for them, they are treated as they should be. If, however, the shopkeeper is robbed by one of the customers and consequently develops an antipathy toward the customers as a group, the motive for refraining from cheating them has been taken away. If, on the other hand, the shopkeeper is motivated by the good will, then the reason to refrain from cheating the customers cannot be taken away⁵. Still further support of the claim that Kant is aware of the issue of the relationship between moral motivation and moral action comes when we consider yet another passage, in which Kant contrasts his approach with a “mixed doctrine of morals” (*Gr* 411). There he says that such a ‘mixed doctrine’ which is “put together from incentives of feeling and inclination and also of rational concepts,

⁵ I would contend that the often used example of two people who go to visit a friend in the hospital, one from inclination and one from duty, is more favorable to Kant if we are to emphasize the fact that Kant takes the view that our sentiments are not in our control. If we notice this view in Kant, then we cannot praise or blame the people in the room for their natural sentiments, but simply admire the one who has the natural inclinations toward generosity, and this would be in keeping with passages such as the one in which Kant refers to a natural feeling of sympathy as one that “deserves praise and encouragement but not esteem” (*Gr* 398).

must make the mind waver between motives that cannot be brought under any principle, that can lead only contingently to good and can very often lead to what is evil” (*Gr* 411).

The issue of whether or not Kant himself was explicitly concerned with the practical importance of the stability of the guidance of the good will is not my concern here, (though I have tried to offer several passages that would lend plausibility to the view that he was aware of this issue). What I wish to suggest is simply that the concern argues in favor of a Kantian approach to moral decision making; the moral law, and respect for the law as seen in acting from duty or according to a good will has the practical advantage of providing us with a stable motive for action that is not present if we rely on features such as sentiment. Further, all of Kant’s metaphysical objections to relying on consequences and other contingent concerns as a ground of our moral obligations could also provide us with practical reasons for rejecting such a ground of morality as well.

There is an additional difficulty with attempts to use the consequences of actions as the means of determining the rightness of those actions. Often, we are unable to see the relationship between any given action (and even any given rule or policy whose purpose is to try to bring about a greater balance of positive outcomes) and the effect this action will have on the future. Consider an example from history. We support Iraq in opposition to Iran, and years later we are confronted with Saddam Hussein. Then, in toppling the regime in Iraq we once again see a rise in the power of Iran. So, even if we remove one danger, we may create another, and there is simply no way to know whether this will happen or not. Of course, utilitarianism would attempt to make calculations

based on long term effects of our actions, and not just the short term results of our decisions. However, two points can be made here. The first point is simply an extension of the general point about the unpredictability of the effects of actions. The further out in time we attempt to calculate the effects of our actions, the less likely we are to be accurate in these predictions. We can perform an action in the name of Democracy only to see that very action decreasing the chances of realizing the sought after goal.

What we must remember is that we are in a position of balancing an unknown -- the results of our actions, against a known -- whether our actions would be acceptable when considered from the perspective of the categorical imperative. Consider the decision to assassinate a world leader for the purpose of overthrowing a government that is viewed as corrupt, even evil. It seems clear that assassination would be considered clearly wrong from a Kantian perspective. It no doubt treats the person who is assassinated as a means to some other end, and we could not will a universal maxim of assassinating leaders who were viewed by others as being corrupt. Any attempt to justify the assassination would thus have to rest on a concern for consequences. But, the calculation of consequences is no more than a gamble, and we are back in the position of arguing that the ends justify the means, even if the ends are not realized, a view that I consider to be untenable.

This issue leads to the second problem with our inability to know the long term consequences of our actions, which stems from the irony inherent in acting in order to bring about a moral world. I would argue that implicit in Kant is the view that we cannot make progress towards a moral world by attempting to calculate the actions that are

mostly likely to contribute to this goal. Quite to the contrary, Kant argues that it is only by always acting in a way that respects the inherent dignity of rational agents and maximizes the freedom of these agents that we can make progress toward a more moral world. Whenever we engage in actions that undermine freedom and dignity for the purposes of ‘making the world a better place,’ we not only have no guarantee that our actions in fact will make the world a better place, we also knowingly undermine the constituent features of our goal: the moral world.

This point is only important if we take the goal of moral action to be progress toward a more moral state of affairs in the world. If our only concern is preventing tragedies from occurring, then it may be that, from a practical point of view, a consequentialist approach is just as good as a Kantian one. However, if we instead take the view that our actions should contribute to a more moral world neither consequentialism nor a Kantian approach can *guarantee* such a moral progress. Thus, in the absence of the ability to guarantee that our actions will result in progress toward a more moral world, we must find another criterion for what it is that makes our actions morally correct. Consequentialism cannot, by definition, do this, since its very measure of success is in terms of the outcomes of our actions, but Kant does offer a separate standard for the morality of our actions: conformity with the moral law. And, as I have tried to show, Kant’s view is that if we act out of respect for the moral law and also work to bring about conditions under which others may do the same and under which the freedom of all can be more fully realized, then we can make progress toward a more moral world.

Herein is the beauty of Kant's concerns about the felicity of sentiments and inclinations. Kant emphasizes the problems that are created if an inclination to perform a duty is taken away from us. This is different than the concern that our inclinations often guide us wrongly (though Kant does suggest that this can happen as well). Rather, any practical concern about inclinations that we can ascribe to Kant is founded on the concern that these inclinations will fail to guide us *in the future*. So, the practical benefit of the good will is that it implies a constant respect for duty, and such a view speaks to the idea that morality is a project, and not a matter of individual or isolated actions. This same theme can be seen another way, when we consider temptations we often feel to make exceptions to moral rules. While exceptions may bring about beneficial short term results, they impede rather than improve our prospects for moral progress. This issue of how we are best to ensure progress toward a more moral world becomes clearer when we consider the issue of making exceptions to existing rules or laws, which may serve as an analogy to cases in which a person wishes to make an exception to the moral law, such as the case of Captain Vere in Melville's *Billy Budd, Sailor*.

Billy Budd and the Importance of Lawfulness

Melville's *Billy Budd, Sailor* is sometimes used as a fictional example that highlights some of the possible problems with a Kantian approach to ethics⁶. In the story, the character Billy Budd is the embodiment of virtue, who is put to death for striking and causing the death of his commanding officer, Master at Arms Claggart, who is as evil as Billy is good. There are many mitigating circumstances, such as the fact that Budd's

⁶ I am thinking in particular of Michael Brannigan's use of the story in his textbook *Ethics Across Cultures: An Introductory Text with Readings*. McGraw Hill, 2004.

outburst was caused in part by Claggart's false accusations that Billy was the leader of a plot to mutiny the ship on which he was impressed. Further, Billy was physically unable to reply with words due to a stutter that rendered him mute when under great stress. However, the ship's captain, Captain Vere, argues to the collection of officers he has convened to decide Budd's fate that "irrespective of the provocation to the blow, a martial court must needs in the present case confine its attention to the blow's consequence, which consequence justly is to be deemed not otherwise than the striker's deed" (Melville 107). The penalty is thus proscribed by martial law, and Vere argues to his officers that they have an absolute duty to follow this law despite the fact that both Vere and the officers feel that Budd is "innocent before God" (Melville 110). The depiction is thus one in which allegiance to the law, albeit formal military law and not the moral law, is given precedence over the pangs of one's conscience and the result is that an innocent is sacrificed before the altar of duty⁷. The question that inevitably is raised by such a depiction is whether Vere's allegiance to duty, while leading to a tragic result, was nonetheless justifiable.

I set aside here all considerations about what Vere's maxim may have been or about the difference between the moral law and martial law. Doing so allows us to see the criticism of Kant that might lay behind any criticism of Vere: adherence to the dictates of the law blinds us to compassion. However, for now I would like to simply

⁷ The novel is actually somewhat ambiguous as to Vere's actual motivation. While he frequently alludes to being motivated by his duty, he also argues that they cannot mitigate the penalty in part because the crew, knowing the proscribed penalty is death, might mutiny at the appearance of arbitrariness (Melville 112-113). For the sake of the argument, however, I will assume his motives were indeed governed by a desire to adhere to the prescriptions of the law.

suggest a defense of Vere's actions. While the novella follows Billy Budd's death with an account of Captian Vere years later, haunted by his decision and, on his death bed, muttering "Billy Budd, Billy Budd" (Melville 129), a movie based on the book, *Billy Budd*⁸, ends differently. The very last lines of the movie are the following: "Men are perishable things, but justice will live as long as the human heart, and the law as long as the human mind." These lines suggest a possible reason to be sympathetic to Vere's decision. Making exceptions to the law undercuts the force of the law as such. So, while tragedies may occur when we refuse to make exceptions to the law, we must weigh these against the harm we risk doing to the law itself if we make such exceptions.

The example of Billy Budd is imperfect, since it refers to actual law, and not to a command of the moral law, but Kant clearly employs an analogous line of reasoning when he discusses our duty of truthfulness in his essay, *On a Supposed Right to Lie from Philanthropy*. There he says that "truthfulness is a duty that must be regarded as the basis of all duties to be grounded on contract, the law of which is made uncertain and useless if even the least exception to it is admitted" (*Supposed Right* 427). The force of law indeed is dependent on the fact that it be followed without exception. In fact, making such exceptions creates a situation of arbitrariness, and justice itself is undermined.

This is not to say that Kant is basing his reason for rejecting a maxim of lying on the basis of its consequences. Kant's concern in this particular treatise is about the law,

⁸ *Billy Budd*. Dir. Peter Ustinov. Perf. Robert Ryan, Peter Ustinov, Terence Stamp. 1962, Warner Brothers.

as he is discussing the ‘right’ to lie in terms of a juridical right⁹. Kant of course gives several rationales for rejecting untruthfulness in his work,¹⁰ so, it should not surprise us that Kant does not discuss the moral law in this essay. However, there is nothing inconsistent in holding that a lie is forbidden because it runs contrary to the dictates of the categorical imperative and also holding that this same lie would also have the result of undermining the force of law. It is important to remember that, just because Kant does not base the morality of actions on a concern for their consequences, this does not mean that the effects of our actions are irrelevant.

The general point here is clear: when faced with a choice between adhering to the law and allowing a tragedy to result, or violating the law so as to prevent the tragedy, there is reason to defend Kant’s allegiance to the law on the grounds that the alternative would be harmful as well, though it would be the law rather than an individual that is harmed. This is not merely a logical point either. When we make exceptions to a law we create a situation in which the law is not a law at all. This point is perhaps most famously made in Plato’s *Crito*. There Socrates adopts the perspective of the laws and the state and speaks for them, saying:

⁹ This fact also seems to make sense of Kant’s claim that we are accountable for the results of our lie, but not for the results of having told the truth. This claim makes much more sense in the legal sense than it would in a broader ethical context.

¹⁰ In addition to the example of the lying promise in the *Groundwork* and the passage quoted above from the *Supposed Right to Lie*, Kant also considers lying as a violation of a duty to one’s self in the Doctrine of Virtue in the *Metaphysics of Morals*. There the argument is that by lying “a human being throws away, and as it were, annihilates his dignity as a human being” (MM429). This is based on Kant’s claim that lying is “directly opposed to the natural purposiveness of the speaker’s capacity to communicate his thoughts, and thus is a renunciation by the speaker of his personality...” (MM 429). It should be noted that, while this passage seems to support a reading of the categorical imperative that emphasizes teleology in terms of purposiveness, as Paton suggests, I would suggest that such a passage is perfectly consistent with the view of rationality I laid out in chapter two.

Tell me Socrates, what are you intending to do? Do you not by this action you are attempting intend to destroy us, the laws, and indeed the whole city, as far as you are concerned? Or do you think it is possible for a city not to be destroyed if the verdicts of its courts have no force but are nullified and set at naught by private individuals? (*Crito* 50b)¹¹

The point being made in this passage is that in making a decision that a law should not be binding in a particular instance, we render the law impotent, as it can no longer be *law* but only a general guide for actions. This occurs whether we attempt to escape the force of law, as Socrates argues an escape from prison would do, or when we simply make an exception to the law. And the foundation of civil society is, at least in part, in its lawfulness. Kant himself recognizes this point as well.

Although Kant does not claim that this is the *ground* for the rejection of maxims such as a maxim of lying to aid a friend, Kant does recognize the line of reasoning I am advancing here as an important consideration. Kant says that lying “always harms another, even if not another individual, nevertheless humanity generally, inasmuch as it makes the source of right unusable” (*Supposed Right* 427). Lying, according to Kant, ‘makes the source of right unusable’ because “truthfulness is a duty that must be regarded as the basis for all duties to be grounded on contract, the law of which is made uncertain and useless if even the least exception to it is admitted” (*Supposed Right* 427). Here Kant makes the point that the practice of verbal commitments rests on the supposition of the

¹¹ Although it is an issue beyond the scope of this dissertation, it is worth clarifying that this is not an argument that civil disobedience is never justified. Socrates is in the position of wondering, along with Crito, whether it would be just to escape the punishment for his practice of philosophy, so the issue is not whether one is justified in breaking the law but whether one is justified in trying to escape punishment for having broken it. Thinkers such as Thoreau and Martin Luther King Jr. do well to show the difference between these two aims when they willingly embrace jail sentences as a consequence of having broken the law.

truthfulness of those verbal commitments. When we sanction an exception to this assumption of veracity, we undermine the very practice of promises or contracts.

Thus, the situation is falsely depicted when Korsgaard suggests that, in the example of lying to the murderer at the door that Kant claims that “I have done my part by telling the truth and the bad results are not my responsibility” (Korsgaard 150). A defense of a Kantian allegiance to the law, whether legal, martial, or moral, is not synonymous to an indifference to the results of following the law. In fact, it is often because we may come to see a given law or practice as resulting in a tragedy, but nonetheless we also recognize the importance of respecting the forcefulness of law as such, that we seek reform.

It seems then that it is often that we are placed in positions in which we feel torn between allowing a harm to occur to a person with whom we are sympathetic and a desire to adhere to policies, rules, or laws so that their force is not undermined. And, when we fall on the side of refusing to make exceptions, it is possible to feel that the preservation of lawfulness is an unsatisfying consolation when we feel that a person has been harmed. However, it is important to note something else that occurs when we feel that adherence to existing policies, rules, or laws results in tragic consequences. When we see such tragedies occur, we are moved to change the conditions that gave rise to them. It may be that we wind up revising a policy or law so that it yields results that are less problematic. So, by staying true to the spirit of lawfulness, we are thus moved to make improvements that we would perhaps otherwise not have made if we simply made exceptions. In a

recent conversation, a colleague, Robert Ladenson, of the Illinois Institute of Technology suggested that such cases can actually be solved another way as well.

As an adjudicator of special education claims for many years, Professor Ladenson commented on having increasingly felt the tension between following the prescription of a law and wishing he could set the law's commands aside, when he thought they would yield a result that he believed to be harmful and even unjust. Ladenson commented that he recognized the fact that one can not place one's own judgment above the law for various reasons, including the practical concern that any judgment contrary to the law simply would not stand up upon review. But he also felt the full weight, like Vere, of allowing the law to produce consequences that were harmful and seemed unjust.

Ladenson suggested that the only true avenue left open in such cases is to recuse oneself from the case, or even retire from a position that constantly put you in a position of having to make such choices. Then, no longer bound by the law by position, one could advocate for changes in it.

Of course, we are not always able to remove ourselves from situations when we feel such a tension between a respect for lawfulness and a concern for avoiding tragic results. One thing that makes Vere's situation compelling is that he cannot relinquish his command, or delay the trial and sentence of Budd, since they are out at sea, in war time, with battle looming close by. And, even when we can remove ourselves from situations in which we feel that a respect for lawfulness would lead to a tragic result one could argue that we take the 'easy way out' if we simply decide to leave that decision to another. There are indeed cases when we cannot help but choose between adherence to

lawfulness and making an exception to the law to prevent an outcome we deem tragic. It seems difficult to suggest that we ought to let the tragedy occur if we feel that a simple compromise of our adherence to lawfulness would be all that was required to avert the tragedy¹². However, this claim does in fact become more plausible when we think in terms of moral progress. When adherence to the law in fact leads to a tragic result, our only solace is that it is this very tragedy that pushes us to find ways to remove such conflicts in the future, while leaving the force of law intact. To see this point more clearly, we turn next to the issue of health care.

Health Care, Lying for Patients, and Reform

A study done in 1999 reported that many “physicians sanction the use of deception to secure third-party payers’ approval of medically indicated care” (Freeman et. al.). One of the many tragedies of the current health care system is that many physicians “have experienced pressure from managed care organizations (MCOs) to reduce utilization and curtail costs” (Freeman et. al.). So, doctors often find themselves in a position of feeling as if they need to choose between allowing their patients to suffer and using deception to help make sure their patients receive care. I should make it clear here that in what follows it may sound as if such deception is clearly wrong, however, for now, I wish to make the slightly milder claim and simply suggest some of considerations that would act in favor of a Kantian view, in opposition to such deception. There are two

¹² What makes the situation of Billy Budd compelling and different from the situation of torture, is that in the example of torture, the morally questionable act – engaging in torture – clearly robs one human being of dignity in the attempt to save others. But it is incorrect to say that the decision as to whether or not to torture is not the only factor that impacts the chance that the lives of others are indeed saved. The case of Billy Budd is more difficult because there is no way to say that the tragic outcome was beyond our control.

considerations we should keep in mind that might argue in favor of a Kantian approach to such circumstances. First, it is worth noting that deception may actually make true reform of a broken system more difficult. Second, the choice between allowing a patient to suffer and engaging in deception may be a false dichotomy, and other options may be open to us that would not entail a violation of any duty akin to Kant's claim that we have a duty to be truthful.

When physicians use deception to help their patients receive treatment, the individual patient is benefited.¹³ However, such actions create the false impression that the system works as it is and is not in need of reform. If, on the other hand, medical professionals did not engage in such deception and patients suffered as a result, this suffering would provide a significant motivation for reform. Now, I do not want to go so far as to suggest that physicians and other medical professionals should have health care reform as a higher priority than care for their individual patients. I am simply offering the view that it is this systematic reform that reflects true moral progress and that an 'ad hoc' approach to injustices may prevent individual harms from occurring, but they offer less hope of true moral progress. The choice between these two alternatives is admittedly a difficult one, and again, I want to stress that I am simply trying to suggest some plausible defenses for a Kantian kind of rigorism in the face of tragedies. However, it may be that such difficult choices are some times more apparent than real.

¹³ However the benefit may be one in need of some qualification. As Werner et. al. point out, such deception can lead to problematic unintended consequences. For example, they point out that when "physicians fabricate medical information for insurers, the information becomes part of their patient's medical records. Not only can this mistakenly impact the decisions future physicians make about patients; it can directly impact patients if they are unaware of the misrepresentation and believe that the chart lore is the truth about their condition" (Werner et. al., 57)

Another study that measured physician willingness to use deception of insurance companies to help procure care for patients added the variable of asking about the option of appealing the insurance company's decision not to provide care for a given patient.¹⁴ In addition to asking physicians to choose between misrepresenting symptoms and accepting the insurance company's decision to deny treatment, the study asked physicians if they felt that they had enough time to appeal the insurance company's decision to deny treatment. The study found that "it is clear that at times some physicians will choose the option that is less hassle: deceiving the insurer" (Werner et. al.). So, the physicians were not always condoning deception in the belief that it was the *only* way to insure patient care, but in some cases the physicians opted for deception instead because they felt it was the more *expedient* way of procuring payments on behalf of their patients. These findings suggest that, at least in some cases, the choice is not truly between deception and allowing a tragedy to occur. There are other options open to us, such as, in the case of physicians dealing with insurers, filing an appeal of a given decision. Such an option may allow a person to avoid deception while also using other, less morally objectionable, methods of helping their patient. This is an important point: often it is tempting to assume that it is *only* by violating a dictate of the moral law that we can prevent a tragedy from occurring. But this assumption is a dangerous one as it can lead us to abandon moral rules too easily. In fact, I would suggest that there is something to be said for the view that adherence to the moral law forces us to become creative in finding ways to

¹⁴ The study (Werner et. al.) actually measured both the general public's willingness to sanction deception and that of physicians.

make a more moral world possible. Further, I want to suggest that Kant himself hints at this point in a few places.

In Chapter Three I made the point that Kant understands nature, guided by a divine author, to be operating in a way that helps to make a moral world possible. However, it is important to look more closely at the way he sees nature as working toward a more moral world. In doing so we will see that Kant in fact claims that it is crisis and tragedy that is largely responsible for forcing us to *ourselves* act to build this moral world. We see this point emerge most clearly in *Toward Perpetual Peace*, particularly, and ironically, in the role Kant assigns to the importance of war.

It is interesting to note that the section of *Perpetual Peace* that Kant entitles “On the Guarantee of Perpetual Peace” is largely devoted to the subject of war. Referring to the providence of nature, Kant tells us that one of the chief ways in which nature operates so as to make perpetual peace possible is that “by *war* it has driven them [the peoples of the earth] everywhere...[and]by war has compelled them to enter into more or less lawful relations” (*Perpetual Peace* 363). He later goes on to discuss the ways that war and discord in general help force people into peaceable relations with one another. He says, for example, that even “if people were not forced by internal discord to submit to the constraint of public laws, war would still force them from without to do so” (*Perpetual Peace* 366). This occurs, according to Kant, because the need to protect oneself from an outside threat forces people to “form internally into a *state* in order to be armed as a *power* against it” (*Perpetual Peace* 366). The interesting point here is that nature does

not itself bring about peace. Instead it brings about conflict and war, and these conditions force human beings to work to overcome these threats to their own survival.

Kant also extends this argument to make the point that even the basest of human inclinations, such as the tendency toward war, which he says “seems to be engrafted onto human nature” (*Perpetual Peace* 365), help serve the end of making moral progress possible. He notes, when talking about the possibility of the forming of a republican constitution, that it is in fact *because* of the self-seeking inclinations within human beings that such a constitution becomes possible. He tells us that the founding of such a constitution is a matter of “arranging those forces of nature in opposition to one another in such a way that one checks the destructive effect of the other or cancels it, so that the ...human being is constrained to become a good citizen even if not a morally good human being” (*Perpetual Peace* 366). According to Kant then, “nature guarantees perpetual peace through the mechanism of human inclination itself” (*Perpetual Peace* 368). So, whether an obstacle to moral progress occurs as a result of our own inclinations or for another reason, we see that these very obstacles provide opportunities for progress toward a state of ‘perpetual peace’.

While this perpetual peace is a political state of affairs and not yet a moral world, we can see that the two share much in common. We can see above that progress toward perpetual peace includes assembling into states, obeying the law, and becoming good citizens. Here we have the essential elements of one piece of the moral world. In the moral world we are all maximally free agents, existing within a community that not only does limit but also works to increase our freedom. It is also a world in which all

members are treated as ends in themselves. Obedience to law and ‘good citizenship’ require at least a respect for the freedom of others. While this is not yet a fully moral world (for which we need to act on moral motives), we can see an argument here as to how progress toward that world can take place, if only by way of analogy.

The discussion Kant gives in *Perpetual Peace* can be seen as offering the view that discord, and by extension obstacles in general, force us to find ways to work together to alleviate these obstacles. I would suggest that we can extend this line of reasoning to cases in which we feel like an existing system or rule leads to a tragic result. Rather than violating the moral law by, for example, engaging in an act of deception as in the case of the physicians lying to benefit their patients, the very experience of feeling the temptation to violate the moral law out of a desire to avoid a particular result can force us to find ways to improve the system. This is not to suggest that *violating* the moral law can lead to moral progress. I want to suggest the opposite conclusion here. It is in *feeling the desire* to violate the moral law that we are presented with a choice. We may either violate the law because we feel that adherence to its dictates will allow a tragedy to occur, or we can force ourselves to ask: How can we avoid a tragic result *without* violating the moral law?

One mistake we often make in choosing to engage in behavior we would normally think of as unethical is that we take the view that there is no other option. But, I would suggest that it is the very act of positing a violation of the moral law as an option that inhibits us from seeing or from taking seriously other possible options. For example, we saw in the case of physicians willing to lie on an insurance form, that such deception was

often chosen out of regard for expediency not necessity. Similarly, in creating an ad-hoc ‘solution’ to a situation felt to be unjust, we build deception into the system rather than working for reform. In taking the idea of engaging in deception ‘off the table’ we are thus more likely to experience reform as the only viable path, and thus work toward it. This is the value of *Perpetual Peace* for our present discussion: the experience of apparent obstacles toward moral progress may also provide us with the very means of making such progress. However, the means to this progress can be impeded, rather than furthered by, violations of the moral law.

The three above examples, of refraining from torture, of following the law and executing Billy Budd, and of refraining from lying to help patients receive medical care, are designed to each offer a ‘line of defense’ for an approach such as Kant’s, where we follow the law or a moral principle, even though a concern for consequences tempts us away from doing so. In the case of torture we see the argument that we are far less in control of consequences than is often depicted, while the moral law gives us a stable guide to our moral duties. In the case of *Billy Budd* we see an argument that exceptions to the law put the law itself in jeopardy. In the case of lying on behalf of patients, we see the argument that ‘ad hoc’ solutions to tragic situations may be less helpful toward the end of promoting a more moral world than are solutions in which our concern for preventing tragic outcomes prompts more systematic reform.

The above considerations all apply to cases in which we see what I would call a specialness of circumstance. In the case of the ticking bomb scenario, the argument is often made that an injunction against torture simply cannot deal with the situation of a

ticking bomb. In *Billy Budd* we see the argument that a given law simply was not intended to cover a case such as Budd's. And, in the example of lying for patients, the argument might be that the system simply fails when it is applied to certain cases, and so accommodations should be made, even if the making of such accommodations would result in action that would otherwise be clearly wrong. The case of the murderer at the door is also an example of a situation in which we may feel a specialness of circumstance. Critics of Kant's insistence that we may not lie do not base their objections on a claim that the person hiding in our home is a friend, but on the basis of the fact that the lie will cause someone presumably innocent to be harmed. So, this type of situation turns into the question about whether we ought to let concern for consequences trump adherence to the moral law.¹⁵ However, the example of Budd also points to the other kind of particularity that would cause us to resist following a universal rule or law. Rather than a specialness of circumstance, we can also see what a claim that Budd himself is special might entail. Budd is the embodiment of virtue, and further, in the novel it is clear that Captain Vere takes on a role very close to that of a father to Budd. So, we have a case where our feelings for an individual or our particular relationship to him or her cause in us a desire to make an exception to a rule or law on his or her behalf. This example can be associated with an approach that is often associated with feminist ethics, or an ethics of care, where a contrast is drawn between impartial

¹⁵ There is also another common approach to the murderer at the door example, which is to try to find room for a lie within Kant, by formulating maxims a certain way, etc. Korsgaard, as I have already discussed at length, takes this approach, as does John Atwell, in *Ends and Principles in Kant*. The difference between Atwell and Korsgaard is that Atwell's approach focuses on a particular view of maxim formulation and Korsgaard's approach depends on a particular reading of what it is that causes a contradictory maxim in the first place.

rules and an ethic that emphasizes the particularity of relationships.¹⁶ It is important for us then to investigate the nature of this contrast, and to determine exactly when it is that feminist ethics and Kantian ethics stand in opposition to one another.

The Challenge from Feminist Ethics: Partiality vs. Impartiality

Virginia Held portrays both utilitarianism and Kantian thinking as opposed to the approach she characterizes as an ethics of care. Generally speaking, Held's criticism is of the view that universal moral principles should always guide our moral decision making. For example, she writes that when "universal principles endanger relationships, the feminist challenge disputes that the principles should always have priority" (Held 91). Held then goes on to suggest that Kant's thought causes just such a disregard for relationships. Referring to his approach, of taking persons as all equally deserving of respect, she writes that "turning everyone into a liberal individual leaves no one adequately attentive to relationships between persons..." (Held 95). But, this criticism seems to me to be unfounded. Just because Kant takes the view that we are to treat all persons as equally worthy of respect, this does not mean that he is committed to the view that we must treat people equally in every respect.

In Kant's ethics beneficence is a wide duty, and so it is left open how we are to exercise that duty. We do not have a duty to be equally beneficent to all, but simply a duty of beneficence in general. Kant makes it clear that, when it comes to the wide duties "the different situations in which human beings may find themselves" (*MM* 392) may

¹⁶ Virginia Held is one author who best epitomizes this view. For example, see Held's *Ethics of Care: Personal, Political, and Global*, especially pages 91-98.

have a large influence on the extent to which we exercise our duties and to whom we direct them. Of course, this is likely not enough to satisfy the nature of a criticism such as Held's, since Held is not suggesting merely that we *may* treat others differently based on the quality or character of a particular relationship, but that we *ought* to do so.

According to Held then, "when a relationship has moral worth, as a caring relationship between parents and children, or a relation of trust between friends and lovers clearly have, the question of the priority or not of impartiality can arise" (Held 95). And for Held, it is partiality that should win out in such instances. This is because Held takes the view that the recognition of the particular nature of a given relationship is itself morally significant.

Held advocates for the view that we ought to be motivated by a recognition of the particular nature of relationships, rather than by a recognition of some quality that all humans share, such as an inherent worth that is bound up with our rationality. She writes, for example: "What a parent may value in her child may well not be what makes this child like every other, but the very particularity of the child and of the relationships between them, such as that she is the mother of this child and this particular person is her child" (Held 93). Her critique of an approach like Kant's is thus that it places the moral motive in the wrong place, requiring us to be motivated, for example, by the recognition that human beings are all ends in themselves. But, it should be clear that the recognition that human beings are all ends in themselves, and thus all deserving of equal respect, does not commit Kant to the view that people should thus be treated in identical fashions. For example, Held depicts a situation in which a father is struggling between spending

time with his students and with his own child. Her argument is that the father should recognize that the particularity of his relationship with his child should give him reason to prioritize it over a responsibility to his students (in terms of amount of time). She then suggests that Kant would arrive at the opposite conclusion. She writes that “the father concludes in this example that with respect to the time he spends fulfilling both duties, his duty to avoid neglecting his students outweighs his duty to avoid neglecting his child” (Held 98). But what is troubling about Held’s example is that she offers no reason *why* the father makes this decision. She simply asserts that this is the decision he reaches. But there is nothing in Kant that tells us that this is the decision that a person in such a situation would or should reach. (While it may be possible to argue that this lack of definitive guidance as to which choice the father should make is a weakness in Kant’s ethics, it is certainly not the weakness Held is attributing to him.)

It is important to see that there are different senses of ‘treating persons equally’ that can be easily confused. In one sense, the injunction to treat all persons equally can be taken to mean that we cannot make an exception on their behalf. In the other sense it can mean that we are to be completely blind to individual differences when we think about the exercise of our moral obligations. It seems to me that Held wrongly blurs these two together. For example, she writes that moral rules, including Kant’s treatment of the categorical imperative, “recommend that we treat all persons equally, including our children ...and that when we have special skills we should use them for the benefit of all persons equally” (Held 99). This view is perfectly compatible with a Kantian approach so long as the special nature of a relationship is only taken into account when we are

deciding to whom we should direct ourselves when fulfilling our wide duties, such as beneficence. Kant's emphasis on treating people equally is directed instead at two concerns: that we ought to treat all human beings as deserving of respect since they are ends in themselves, and that duty, to be deserving of the name, cannot admit of exceptions. So, the question we should be asking is: What are we to do in situations when the special nature of a relationship is being weighed directly against an injunction that we ought to be able to universalize a maxim, an injunction which takes the making of exceptions as the surest sign of a violation of duty?

The assertion in feminist ethics that we ought to give special weight to the particular nature of a given relationship is less appealing when this would amount to violating a moral rule for the sake of a particular individual. In the examples of *Billy Budd* and of lying to procure medicine for a patient in need, earlier in the chapter, we saw that the decision to make an exception to help someone in need could cause several problems. I suggested there that violating moral rules has the consequence of weakening respect for lawfulness as such. Further, I suggested that a decision to violate a moral rule often falsely supposes that our only choices are violating a moral rule and allowing a tragedy to occur. This point led to the additional point that it may be the case that progress toward a more moral world is made more, not less, possible if we are forced to change rules rather than violate them. When we add the consideration that the reason we seek to make an exception to a moral rule is due to the specific nature of our relationship to a specific person, further complications are added. Most importantly, we need to be

wary of the ways in which special relationships may unwittingly turn into unjustified biases.

Held tells us that special consideration based on sympathies with a white supremacist would not be justified while considerations based on relationships such as family affiliation or friendship are more acceptable (Held 95). But this of course forces us to ask about many instances in which we are unaware of the reasons why we favor one person over another out of a feeling of kinship. One argument in favor of affirmative action is that it compensates not for past instances of racism but for unintentional biases on the part of employers. Setting the affirmative action debate aside, one could see that it is reasonable to suggest that employers and coworkers might unknowingly favor those who look like they do. This reminds us of the problems, pointed out in Chapter Three, of attempting to apportion individual happiness to individual virtue or worthiness to be happy. There I suggested that we should not think of the highest good possible in the world in this way because doing so would suggest that we had a duty to evaluate a person's worthiness to be happy and attempt to promote their happiness accordingly. The first of these requirements is likely impossible, and even if possible, the idea of only promoting the happiness of those we think 'deserve it' is morally problematic.

This issue is exacerbated when we introduce the element of special relationships. Then, our attempt to promote a person's happiness would become based not even on their apparent virtue, but on our fondness for them. There are two obvious problems with this. First, as was pointed out earlier, fondness for a given person is not something that can be commanded, so even if making moral decisions based on the specialness of a relationship

was laudable, there is no guarantee that people would themselves recognize the specialness of a relationship that an author like Held would suggest is obviously special (such as that between parent and child). Second, if the objective nature of a relationship, such as the fact that it is one between friends or one between a parent and child, is not enough to ensure that we subjectively experience the relationship as special, then our basis for experiencing a relationship as special becomes purely subjective and thus arbitrary.

In defense of Held, her position, at least in *Ethics of Care*, is not that we should be partial to particular others in cases when this partiality comes in direct conflict with a moral rule. Instead, she argues for a broader claim, that an acknowledgement of the specialness of particular relationships is an important aspect of our moral lives. I am sympathetic to this view, and I have tried to show that there is nothing in Kant that would prohibit it. I have attempted to show that it is for this reason that we must distinguish between two notions of treating people equally. I have tried to show that it is a mischaracterization of Kant to portray him as advocating a view that we should be blind to individual differences and the special nature of relationships. In this sense, it may be appropriate, even for Kant, to treat people differently. Kant would reject only the more drastic claim, that partiality should take precedence over the moral law when the two come into direct conflict. And, I would suggest that he would have good reasons for rejecting such a claim.

I offer the above considerations not as decisive arguments in favor of Kant's approach to tragic circumstances, but simply as an argument that Kant's approach should

not be dismissed too easily, as Korsgaard's remark that it is "grotesque" would have us do. I have argued that, if moral progress is our goal, then the approaches I have contrasted Kant's with are themselves problematic, and this makes his own approach more plausible. Korsgaard's own approach is however unique in that it recommends that we use multiple standards by which we can evaluate the morality of actions. For Korsgaard, we should use one standard in "everyday life," while a different standard may be necessary for extreme circumstances (Korsgaard 150). With this in mind, we turn now to examine her argument for a 'double level' theory.

Korsgaard's Double-Level Approach

The way that Korsgaard differs from Kant stems from the way each sees the role of the Kingdom of Ends. For Kant, the Kingdom of Ends is both a social goal and a normative principle. This is because we are to try to make the Kingdom of Ends possible, but we are also to act as if we are legislators in such a kingdom already. By contrast, Korsgaard takes the view that in some cases "the Kingdom of Ends can become a goal to seek rather than an ideal to live up to..." (Korsgaard 153). These cases are what she refers to as "evil circumstances" (Korsgaard 153), and she uses the example of the murderer at the door as an example of such a circumstance. For Korsgaard, the reason we may abandon the Formula of Humanity and the Formula of the Kingdom of Ends in such evil circumstances is that they give "implausible answers when we are dealing with the misconduct of others and recalcitrance of nature" (Korsgaard 151). For example, on Korsgaard's reading (with which I have no objection) the Formula of Humanity would clearly prohibit lying to the murderer at the door. For her, this is an example of an

‘implausible answer.’ I have already tried to argue for the view that Kant’s answer may not be as implausible as Korsgaard suggests, and I will return to this issue later.

However, what is important to focus on now is not Korsgaard’s differences with Kant but her differences from others who would reject Kant’s strict adherence to moral principles in “evil circumstances”. While others might take the view that Kant’s ‘implausibility’ in such cases is an argument for the implausibility of his ethics as a whole, this is not Korsgaard’s view.

It is important to remember that Korsgaard argues that the Formula of Universal Law is not equivalent to the other formulations, which is an issue I took up in Chapter One. In particular, her view is that the Formula of Universal Law is less stringent than the Formula of Humanity¹⁷ and thus “provides the point at which morality becomes uncompromising” (Korsgaard 154). So, to summarize Korsgaard’s position, she takes the view that we should guide our conduct by the Formula of Humanity in most situations, but in certain situations we should abandon it and follow what she takes to be the less stringent formula, the Formula of Universal Law. Thus she offers what she calls a double-level theory, in which we are given two sets of moral principles, one for ‘normal’ situations and one for ‘extreme’ situations. What I wish to investigate here is the plausibility of this double-level approach as applied to Kant’s categorical imperative.

¹⁷ She describes the Formula of the Kingdom of Ends as a “corollary” to the Formula of Humanity, and thus the two are either supposedly equivalent or much more closely related than either of the formulations is to the Formula of Universal Law. I have offered my own argument in Chapter Three regarding the relationship between the Formula of Humanity and the Formula of the Kingdom of Ends, but I do not reject her claim that the two are equivalent.

However, in order to do this it will be important to first clarify Korsgaard's discussion of double-level theory, which will include a brief discussion of its roots in Rawls.

A discussion of Rawls' *Theory of Justice*, which is the inspiration for Korsgaard's use of a double-level theory, is beyond the scope of this project. However, Korsgaard makes use of several terms introduced by Rawls, so it is necessary to clarify a few of these terms at the outset. The first important distinction Rawls draws, which Korsgaard makes use of, is the division between ideal theory and nonideal theory. Ideal theory deals with the question of "what a perfectly just society would look like" (Rawls 8). In such a society each person is "presumed to act justly and to do his part in upholding just institutions" (Rawls 8). This is contrasted with nonideal theory, which is used to deal with "the principles that govern how we are to deal with injustice" (Rawls 8). These two terms are related to, but should not be confused with, two other terms, what Rawls calls the "general conception" of justice and the "special conception."

Rawls lays out two basic principles that are intended to govern an ideal society: that all persons have equal rights to "the most extensive basic liberty compatible with a similar liberty for others" (Rawls 62), and that any inequalities in society are arranged so that they benefit all and are "attached to positions and offices open to all" (Rawls 62). He also tells us that these principles are ordered in terms of priority, so that it is impermissible to trade away liberties or equalities of opportunity to allow for their uneven distribution, for example, for the sake of universal economic benefit. Rawls refers to these two principles as a "special case of a more general conception of justice" (Rawls 62). So, the serial ordering of these principles is associated with the "special

conception” of justice. Rawls then contrasts this with the “general conception,” which he expresses the following way: “All social values – liberty and opportunity, income and wealth, and the bases of self-respect – are to be distributed equally unless the unequal distribution of any, or all, of these values is to everyone’s advantage” (Rawls 62). Thus the “special conception” is more restrictive than the “general conception,” since the general conception would allow for trade offs of, for example, equality of opportunity, if they were to benefit all, while the special conception would not allow this. We then have two sets of principles, one more stringent and one less so. This is important because of the question of when we are to use each, which is the question Korsgaard wants to address in dealing with Kant.

According to Rawls, the two principles that comprise the “special conception” of justice “define then a perfectly just scheme; they belong to ideal theory and set an aim to guide the course of social reform” (Rawls 245). But he makes it clear that this conception was not intended to deal with situations of injustice, “so it is possible that [in such cases] they no longer hold” (Rawls 245). So, the suggestion is that this conception may be abandoned in favor of the less stringent general conception as long as in so doing “social conditions are eventually brought about under which a lesser than equal liberty would no longer be accepted” (Rawls 247). This is what Korsgaard refers to as a double-level theory. Korsgaard then argues for the view that we should treat the various formulations of the categorical imperative in the same way Rawls treats the two conceptions of justice. However, in order to see if this suggestion is plausible, it is

important to look at whether we can appropriately draw an analogy between Kant's approach and the sort of ideal theory laid out by Rawls.

Korsgaard suggests that the "standard of conduct he [Kant] sets for us is designed for living in an ideal state of affairs" (Korsgaard 149). It may be helpful then to begin by considering whether this is an accurate depiction of Kant's view. Korsgaard associates Kant's commands with an ideal state of affairs because she notes that, for Kant, "we are always to act as if we were living in a Kingdom of Ends, regardless of possible disastrous results" (Korsgaard 149). However, this depiction of Kant's view is perhaps misleading. Stated this way, the impression could be given that Kant is asking us to suppose that all people already act out of respect for the law and treat others as ends in themselves. But of course, this is not at all what Kant has in mind when he says that "every rational being must act as if he were by his maxims at all times a law giving member of a kingdom of ends" (*Gr* 438). Just because we must follow the moral law, this does not mean that we have no obligation to try to bring about the Kingdom of Ends. If we are to 'act as if we were [already] living in a Kingdom of Ends,' we would lose sight of any command to help bring about such a kingdom. It is our very awareness that we do not yet live in a Kingdom of Ends, or a moral world, that helps us to see our duties to bring it about.

Nigel Harris makes an argument that touches on this point in "Kantian Duties and Immoral Agents." There he suggests that our decisions about what our duties are in a given situation take place in a two stage process. In the first stage, we must include our awareness of the possible irrational motives of persons with whom we are dealing, such as the fact that a person at our door may in fact be intending murder. This awareness

helps us to consider our options. Harris, for example, makes the point that we would never consider lying to a person at our door who was looking for our friend unless we had reason to believe they had nefarious intentions. In the second stage of the decision making process, according to Harris, we bring in the categorical imperative, which bids us to consider the persons with whom we are dealing as rational agents. So, even though we are aware of their immoral intentions, we still must consider the fact that they expect us to tell the truth. Thus Harris concludes that even in this two stage process we still have, according to Kant, a duty not to lie to the murderer. He does little in the article to wrestle with such cases, but his general point is that we would not take into account the fact that a person was irrational if we took Kant's formulation of the kingdom of ends as calling on us to assume that we currently live in a world where all agents are wholly rational¹⁸. Thus, it makes more sense to take Kant's claims that we are to always act as if we were members of a Kingdom of Ends as indicative of Kant's view of how we are to bring such a kingdom about and not as a sign of indifference to the fact that the Kingdom of Ends is not yet a reality. This distinction is important, because some criticisms of Kant's approach, such as one offered by Michael Phillips in his article "Reflections on the Transition from Ideal to Non-Ideal Theory," seem to rest on this misinterpretation.

Phillips refers to Kant's approach as one of "Moral Purism" (Phillips 552), and he levels several criticisms against such a view. But Phillips' criticism of "moral purism" has two distinct sources, and these two sources come from the two ways in which we can understand Phillips' comment that "Moral Purism becomes the view that the *principles* of

¹⁸ See especially pages 340-342.

Ideal Theory ought to govern behavior in the actual world” (Phillips 555). In one sense this comment means simply that Kant, as a “moral purist” would have us act on the same principles we would act on if the world were morally ideal, even though it is not. This is, I think, true of Kant. The other sense of this comment can be taken as meaning that Kant would have us act as if this world existed already. I have suggested that this is however not an accurate portrayal of Kant’s view. With this in mind, it becomes easier to respond to some of the criticisms made by Phillips.

One criticism Phillips brings up against Kant’s “moral purism” is an issue of logical possibility rather than practical implausibility. Since, on Phillips’ reading, Kant “would have us act on a morality that is appropriate to an ideally structured society” (Phillips 556), some of the duties that would exist in such a society may be impossible to perform in a society that is less than ideal. He uses the example of a duty of loyalty to friends, noting that in a society where everyone else is a devoted egoist, it would be impossible to form friendships and thus impossible to carry out a duty to be loyal to them (Phillips 557). But this criticism rests on the misreading of Kant that I mentioned above. For Kant, I do not look at any society, real or imagined, and generate a list of duties based on the image of that society. Rather, I look at the nature of rationality as such and decide what a rational agent can and cannot will¹⁹. The mistake Phillips makes is to confuse two related concepts: acting as if we were members of a kingdom of ends (even though it does not exist), and acting as if the kingdom of ends already existed. The

¹⁹ As we saw in Chapter Two, this will lead to an appreciation for the empirical conditions we are in, since there are duties, for example, that may apply to dependent beings such as human beings, but not self-sufficient beings, and so, given that we are dependent beings, we cannot rationally will, for example, maxims of non beneficence.

former command can be traced to the moral law, while the latter derives duties from a conception of what the ideal state of affairs (the Kingdom of Ends) is like.

This same confusion can be found in another criticism Phillips makes of “moral purism.” This time the claim is that such an approach can lead to morally repugnant conclusions about how we should behave. Phillips notes that we could imagine a morality that envisions as its goal “the creation of a world in which there is universal respect for persons” (Phillips 559), which is a view I would associate with Kant. Regarding this morality Phillips further asks us to imagine that it places a high premium on respect and on truthfulness. So far, this is again in keeping with Kant. However, he then makes a leap and suggests that we should imagine that this morality also “includes candor – as a form of respect -- where we would now avoid it (in the Kingdom of Ends we have far less to fear from our rivals, adversaries and enemies – if there be enemies here)” (Phillips 559). After making this addition, he uses it to argue against the plausibility of such an approach by noting that, if an enemy of my friend was to ask about my friend’s weaknesses and I were to respond candidly, I would have acted unjustly (Phillips 559). While Phillips is right to say that we would act unjustly in being fully candid, he is wrong to suggest that Kant’s command to treat people as ends in themselves or the duty not to lie are equivalent to offering up any sought after information. What causes Phillips to make this leap, however, is the fact that he seems to read Kant as suggesting that we are to act as if we were oblivious to the fact that our candor could be used for immoral purposes. But, as I have tried to show already, this is a misreading of Kant’s approach.

Phillips is on firmer ground when he returns to his more general point that acting “on a principle that is part of one’s moral ideal may impede a wider realization of that ideal” (Phillips 559). Phillips’ basic argument, like Korsgaard’s and many other critics of Kant’s rigorism, is that his position leads to unfeasible and even immoral conclusions. Phillips makes the point that the type of society Kant envisions, as the Kingdom of Ends realized on earth, could not be achieved “without occasionally treating another as a means only” (Phillips 560). He then goes further and suggests that to argue against this claim is “absurd”. He writes as follows:

This is not to say that the end – a greater realization of one’s Ideal – always justifies the means. My point requires only a denial of the absurd claim that the end never does so. If...by one small lie we could bring about a world in which everyone acted as if he/she were members of the Kingdom of Ends, we should be foolish and immoral to refuse (Kant notwithstanding)²⁰. (Phillips 560)

There is not much left to say to critics like Phillips except to simply deny his premise.

This dissertation has been, in large part, an effort to do just that. As Phillips notes, Kant would reject the claim Phillips makes, that the lie would be worth the end attained by it. However, it is important to look again at some reasons why he would do so.

Imagine a person who took the view that he or she ought to tell lies to foster the conditions for more truth²¹. I find this approach to be chilling in part because of its seeming appeal. It seems at first to be *prima facie* the case that an increase in honesty on

²⁰ Phillips seems to suggest that the lie to bring about the Kingdom of Ends would be justified because it is helpful in attaining that goal, he does not claim that the goal is impossible without the lie. This distinction is important because the argument is thus that actions that would be considered immoral in an ideal world might be acceptable in a non-ideal world simply because they facilitate the ease with which such a world can be brought about. But, this claim gives absolute moral priority to the end goal of the moral world and ignores all other facets of moral action. Kant’s insight is that the end and the means cannot be separated.

²¹ Phillips himself uses the example of lying to a McCarthyist Committee about the activities of one’s university, in order to preserve the atmosphere of intellectual honesty present within the university.

the whole is worth one small lie, especially when we feel we are lying to immoral people. However, this rationale is indeed used to justify continual infractions, since the world can always be brought closer to the ideal. Kant's insight is that there is only one plausible place to draw the line when it comes to the question of when, if ever, the ends justify the means, and that is to say they never do. While many argue, as Phillips does, that this is simply absurd, the counter-claim, in my view is more troubling. This is the case because, if we do not say that the end never justifies the means, we are put in the precarious position of never having a clear unequivocal standard for when the end *does* justify the means²².

A further response to the approach taken by Phillips also takes us back to the argument in Chapters Two and Three. There we saw the connection between the goodness of ends and the moral law. Rational nature is the only self-existing end, and the command to treat humanity as an end in itself leads us to try to promote the autonomy that is the defining feature of our rationality, and thus work towards a Kingdom of Ends on earth. However the very thing that makes rational nature an end in itself is the ability to give ourselves the moral law and act accordingly. So, to try to accomplish the creation of a moral world by violating the commands of the moral law is, for Kant at least, self-contradictory. The value of the moral world cannot be separated from the value of the human being as a being acting from a respect for the moral law. And, Kant seems right

²² I would go so far as to suggest that Phillips himself begins to fall down this 'slippery slope.' He goes so far as to argue that "we have no obligation to act on a higher more restrictive moral standard in relation to others than they act upon in relation to us" (Phillips 561). I find this claim deeply troubling. The suggestion here is not that acting on strict moral standards permits immorality to occur, or even that acting on such standards makes a moral world more difficult to attain. Rather, it suggests that it may be permissible to deviate from strict moral standards whenever it disadvantages us. This claim also has the absurd result, if taken literally, of justifying torture to those that would torture us.

on this point. The idea of willing the existence of truth by the means of untruth seems like a perfect example of the type of contradiction that speaks to an attempt to separate the morality of the end from the morality of the means to its attainment.

The above point leads to an important difference between Rawlsian ideal theory and Kant's ethics. In ideal theory as depicted by Rawls, we ask what principles would characterize an ideal society, and we use these principles to think backward and ask how we can bring about a society that best embodies this ideal. But those principles are not themselves guides to the realization of this ideal. In contrast, Kant is concerned with principles that guide our conduct, and, while he argues that it is in fact possible that moral action will bring about a more moral world, he takes the view that these principles hold for us whether or not they lead to an ideal state of affairs. So, we should be careful not to take Kant's "moral purism" as synonymous with a kind of ideal theory if we are to avoid wrongly ascribing many implausible views to him²³. This difference between Rawls and Kant is significant, as it leads back to the question of the source of our moral principles.

Korsgaard's argument, that we ought to use the Formula of Humanity in most circumstances, but employ the Formula of Universal Law in "evil circumstances," rests on her case that the two formulations are non-equivalent. In Chapter One I considered Korsgaard's case for non-equivalence and argued against it. But even if I am correct that the formulations are to be understood as equivalent (at least in the sense that the various

²³ One of the examples Phillips uses is deserving of specific comment. Phillips suggests that codes of ethics such as the American Bar Association's code of legal ethics are written on the assumption of a situation of justice. He goes on to suggest that, in instances where a system such as the legal system contain serious injustices, that we cannot look to these codes to guide our behavior. (See page 552 for this view.) I strongly reject this claim and take the view that departing from such codes in these situations would just as likely inject more immorality into a situation and into the system as a whole as it would to help alleviate injustice.

formulations would not yield results that contradicted one another), we could still imagine someone, perhaps Korsgaard herself, recommending that we follow the categorical imperative for most situations, but that we should develop a separate standard for our conduct in “evil circumstances.” This leads to a second sense in which the issue of the equivalence or non equivalence of the formulations is significant. If the formulations are in fact equivalent and we cannot therefore look to a different formulation of the categorical imperative to guide our conduct in these “evil circumstances,” then whatever standard we would use for such situations would need its own derivation. Otherwise, this second standard of conduct would seem to be an arbitrary line in the sand. To my mind this is the single most important consideration in favor of Kant’s approach and against a double-level approach.

It would in some sense be better to abandon Kant altogether than to propose that we are to follow Kant in trying to found our moral principles on universal and necessary features of rationality...except in some cases. Approaches such as consequentialism or feminist ethics begin by looking at the world, and look backward for principles that can deal with the world effectively. Feminist ethics starts with an appreciation of the particular relationships we have as individuals and asks us to make moral decisions with these relationships in mind. Utilitarianism observes that all people seek well-being and asks us to consider the world around us and to ask which actions will best contribute to the well-being of all. Kant begins in the opposite direction. He begins by trying to find the grounds for the concept of duty as such, and only then examines what our individual duties are. A double-level approach, if both “levels” cannot be derived from the

categorical imperative, splits the difference between these two approaches. It follows Kant in the derivation of the principle that will guide most of our moral lives (the categorical imperative), but it works backwards when it comes to “evil circumstances,” moving from our concern about conditions we encounter in the world and asking us to come up with moral principles that are best situated for dealing with these circumstances. The result is a moral philosophy that has a ‘split-personality.’

Taking an approach such as Korsgaard’s gives the mistaken impression that the reason Kant’s categorical imperative is acceptable in most circumstances is because it yields “acceptable results,” but such a calculation is of course anathema to Kant’s entire way of thinking about ethics²⁴. Kant makes it clear that “the facility with which a principle can be used and its apparent adequacy furnish no quite certain proof of its correctness but, instead, awaken a certain bias against rigorously investigating and weighing it in itself and without any regard for what follows from it” (*Gr* 392). One might argue coherently that concerns for consequences, whether in general or to particular individuals, should guide our conduct. And one could argue also that these concerns must give deference to an independent standard of our moral conduct, derived from an investigation of practical reason. But, what seems difficult to argue is that we ought to do both, that is, that we should base our notion of moral obligation on practical reason, except when it yields undesirable results, and then we should supplement it with a

²⁴ Korsgaard argues that both of her standards for conduct are derived from the categorical imperative, and thus she has not strayed from an approach that grounds our obligations in considerations of rationality. However, the basis for her argument that we ought to use the less stringent formula in “evil circumstances” is nonetheless based on a concern for the results of the formula’s application. For example, she writes “we can say that the Formula of Humanity is inapplicable because it is not designed for use when dealing with evil” (Korsgaard 151). Also, she writes of the same formula that it “gives implausible answers...” (Korsgaard 151).

concern for dealing with those results. But even if this approach were plausible, I am unconvinced that it would be particularly helpful in practice.

We should remember that Korsgaard offers us two standards for conduct: the Formula of Humanity and the Formula of Universal Law, which she argues yield different results in practice. To defend this view she applied both formulations to the case of the murderer at the door, and derived different results of that case using the different formulas. While I have already argued against her treatment of this case under the Formula of Universal Law, we might still ask about other cases, and ask how often it would actually be the case that, even on Korsgaard's interpretation of the Formula of Universal Law, that the two formulations would yield different results. It is helpful to remember that Korsgaard based her argument, that the Formula of Universal Law would allow us to lie to the murderer at the door, on three important provisions: First, we are aware of the murderer's intentions. Second, the murderer does not know that we are aware of his intentions. So, third the murderer would have no reason to believe we are lying to him. Thus, even if everyone were to lie in these situations, such lies would never be uncovered and the practice would not undermine itself²⁵. We could ask then if this same way of thinking would, for example, allow us to lie to insurance companies to help patients. It seems as if the answer is that it would not. On Korsgaard's examination of the murderer at the door, our deception could be universalized because of the other two provisions built into the case, but these provisions are not present in a case such as lying

²⁵ This further presupposes Korsgaard's use of what she called the "Practical Contradiction Interpretation" of why it is that maxims cannot be universalized, which I argued against in Chapter One. However, for our purposes here, I will simply make use of her interpretation.

to deceive insurance companies. If physicians universally lied on insurance forms, this practice would undoubtedly undermine itself, and the insurance companies might, for example, adjust their policies so as to account for this deception²⁶. So, either formulation of the categorical imperative, even on Korsgaard's own treatment, would yield the same result here. In fact, one regrettable omission from Korsgaard's analysis is a treatment of this question: Of the circumstances we might call "evil," how many would be handled differently by the two different formulations of the categorical imperative, even if we were to accept her arguments for their inequality?

One further concern about Korsgaard's argument that we should use a less stringent moral principle in dealing with "evil circumstances" is that she does not give us much sense of which sort of circumstances are to count as evil. Nor does she offer an account of a process by which we can determine this. She does say that we must do better than the "common-sense approach" (Korsgaard 150), which simply tells us to "depart from our ordinary rules and standards of our conduct when the consequences of our following them would be 'very bad' (Korsgaard 150). She correctly points out that

²⁶ Korsgaard offers an analogous example in Chapter Three of *Creating the Kingdom of Ends*. There she puts the issue the following way: "The method of dealing with natural acts which I have just suggested focuses on the question of whether you could readily achieve your purpose ... in a world where your action was the universal method for achieving that purpose" (Korsgaard 99). She uses the example of murdering one's competitors in order to get a job in order to illustrate this process. My goal of getting a job through violence would be thwarted in either of two possible ways. First, "[t]he use of violent natural means for achieving ends cannot be universalized because that would leave us insecure in the possession of these goods, and without that security these goods are no good to us at all" (99). If that is the case, then a universal practice of killing to get jobs puts our own job (and our life) in persistent danger and my maxim undermines itself. The second way in which Korsgaard suggests that the universal practice of violence to get jobs undermines itself is to "consider whether the social conditions that allow violence to work as a method of achieving this purpose would exist if it were the universal method" (99). Her point here is fairly straightforward: if everyone used violence to get jobs, employers would adjust the hiring process so as to undermine the effectiveness of the tactic. Korsgaard suggests some possibilities: "Perhaps no one would be told who the candidates were, or people would even keep it a secret what job they held" (100).

such an approach leads down a “consequentialist slippery slope” (Korsgaard 150) where the temptation is to justify incrementally greater violations of moral standards in favor of incrementally increasing concerns for consequences. Her alternative is thus to turn back to the analogy with Rawls’ double-level approach. But this comparison is not quite appropriate because Rawls does give us guidelines as to when his “general conception” of justice should guide us and when the more restrictive “special conception” is adequate for the task. This is because we are guided by an idea of the goal we are striving toward rather than by a conception of the commands of the moral law. So for Rawls it makes sense to argue that temporary inequality may be necessary (in the form, for instance of affirmative action) to bring about greater equality²⁷.

It seems to me then that Korsgaard’s double-level approach has several problems. In the first place, Korsgaard bases her argument for this approach on the claim that the formulas of the categorical imperative are not equivalent, which is a claim I argued against in Chapter One. If Korsgaard is incorrect about the issue of nonequivalence, we are presented with a second question of whether we should simply posit another moral principle to deal with situations in which we feel that our usual moral principle (the categorical imperative) yields unacceptable results. But this causes two problems. First, it causes us to ask about the source and justification of the principle we have developed for use in these special circumstances. Second, it jars against the spirit of Kant’s ethics in that it puts the concern for the consequences of the application of a moral principle before a concern for whether the principle rests on a legitimate foundation. In addition, we have

²⁷ The example here is actually Korsgaard’s. See page 148 of *Creating the Kingdom of Ends*.

the practical concerns of whether a second principle would in fact yield results different from the principle to which we normally appeal when making our moral decisions. And finally, we would need criteria to know when to use which principle. For all these reasons, while a double-level theory may be plausible in some respects, such as in Rawls' project, we ought to reject this approach if we are to retain the essence of Kant's thinking. And this conclusion leads to a further conclusion, that Kant's approach to ethics, even in the face of situations we may feel as tragic, can be plausibly defended against various alternatives.

CHAPTER SIX

CONCLUSION - THE PLACE OF HOPE IN AN IMPERFECT WORLD

This dissertation is an attempt to address two issues that are often associated with Kant's ethics. The first is an interpretive concern, which I believe wrongly portrays Kant as someone who is indifferent to teleological concerns and lets an allegiance to the dictates of rationality blind him to all other concerns. The second issue is related to the first, since it represents the view that, regardless of whether or not Kant allows for teleological concerns in his thinking, his focus on strict adherence to the moral law simply is implausible in situations in which this adherence may be seen to lead to tragedies and even injustices. The first three chapters of this dissertation have attempted to deal with the first concern, and the final chapter is intended to provide a response to the latter.

In Chapter One I attempted to show that the avenue Christine Korsgaard attempts to open for us, of separating the formulations of the categorical imperative so that Kant can give us two guidelines for action instead of one, is not in fact in keeping with Kant's own view. Korsgaard's ultimate motivation for making this divide was the worry that the strictness of the Formula of Humanity restricts our own actions so much that it puts us in the position of allowing tragedies and injustices to occur out of a misplaced over-allegiance to a particular notion of rationality. In Chapter Two I argued that Kant's notion of rationality has a legitimate teleological component, so that allegiance to the

moral law can be seen as inseparable from a concern for certain ends. This is important because it helps us to see that, when we refuse to violate the moral law, we do so not simply because we value rationality for its own sake but because we value the autonomy of ourselves and others. In Chapter Three I attempted to demonstrate that Kant saw the Kingdom of Ends on earth as the ultimate end of moral action and as possible in this world. This step in the argument helped us to further see that Kant is not indifferent to the cause of moral progress. Indeed, Kant sees allegiance to the moral law as not only compatible with moral progress but as the only legitimate means to it. In the final chapter I attempted to defend this position, arguing against other approaches to dealing with tragic circumstances, especially as it pertains to the goal of making a more moral world possible. This chapter culminated in a return to a discussion of Christine Korsgaard's treatment of Kant. However, the issue in Chapter One was interpretive, and the question was whether Kant would accept the claim that the formulas of the categorical imperative are non-equivalent. The issue in Chapter Four concerned Korsgaard's other claim, that, regardless of what Kant himself might have said, we *should* separate the formulas of the categorical imperative, allowing us to have one standard of morality for ideal circumstances and one for non-ideal circumstances. Korsgaard uses Rawls' double-level theory as analogous to her approach to Kant, and I argued against the plausibility of such an analogy.

Ultimately, Kant's approach to morality, and my defense of it, cannot help but feel unsatisfying to some. This is because Kant makes it clear that the commands of the

moral law will not necessarily make a perfectly moral world a reality. All we are entitled to do is to hope for such a state of affairs. Additionally, nothing in this dissertation amounts to an outright denial of Korsgaard's claim that adherence to the moral law will sometimes allow a tragedy to occur. However, what I have tried to suggest is that Kant can still speak to this concern. First, Kant shows us that tragedy, even injustice, is an undeniable feature of human life, one we are not entirely able to control. What we can control is our own actions, and we must separate the question of whether our own actions are moral from the question of what is accomplished through them, since these outcomes are not as much in our control as we might like to believe. Further, Kant is able to show a connection between the grounds of morality and the goal of a moral world in a way that other approaches to morality cannot.

I have also tried to argue, primarily in the final chapter, that there are some practical responses to the worry that allegiance to the moral law cannot be reconciled with a desire to prevent tragedies from occurring. In many occasions this is a false dichotomy. Frequently commentators make the mistake of making logical leaps in their portrayal of Kant's depiction of such conflicts. For example, the command that we may not lie to the murderer at the door does not imply that we must simply escort the murderer to our friend. Similarly, many times we tend to portray situations in our lives as containing this same sort of set of choices: either we do something we would normally count as immoral, or a tragedy occurs. However, I have tried to show that this portrayal is often misleading. The depiction of situations as only offering two alternatives in fact often paves the path for a decreasing respect for the commands of the moral law. This is

so because we can grasp the concreteness of a tragic outcome. A person dies. A person does not get necessary medical treatment – these things are real to us. In contrast, the reality of the moral law is much more difficult to grasp.¹ Since the reality of the moral law is less clear to us, the temptation will always be to prioritize the more concrete reality of a given situation, and thus we are constantly at risk of diminishing the importance of the moral law. This by itself is a danger, but portraying situations as having only two choices also limits our moral imagination. We too easily depict the situation as between a concrete harm and an abstract dictate of rationality, and conclude the latter simply should give way. But, if we were to instead reason that this dictate of rationality must never be sacrificed, we would be pushed to look for alternative ways to confront a given situation.

I take the view that when we dismiss thinkers such as Kant, who will not allow us to be flexible in our moral obligations in times like these, we run the serious risk of allowing ourselves to behave increasingly immorally under the belief that we act so as to bring about a state of greater morality in the future. So, my hope is that perhaps Kant can act as a heuristic device against this tendency, if nothing else. And I think there are several examples from history that show us how important it can be to respond to tragedy and injustice by remaining true to moral ideals.

There are many examples from history of those who refused to practice injustice or engage in immoral acts, on the belief that the injustice of the situation in which they

¹ It is well beyond the scope of this project, but there is an analogy to Plato's discussion of the importance of the belief in the reality of eternal ideas, especially of the good, as both difficult to grasp and essential to valuing the just life, especially as it appears in the *Republic*.

found themselves was enough to justify actions that they would ordinarily never find permissible. Most often, these examples portray unique individuals, who we think of as inspiring figures but as embodying a morality that is profound but unrealistic for ordinary people. Examples of such people include Gandhi, Martin Luther King Jr., and Nelson Mandela. But of course it should be remembered that, while these individuals were prominent leaders, they each led movements in which many ‘ordinary’ people acted in extraordinary ways in the face of tragic and unjust situations. And there is one example in particular that stands out to me as an example of what it is possible for people to accomplish without engaging in actions that would otherwise clearly be wrong.

A year or so ago, a colleague of mine named Richard Middleton-Kaplan brought the book *We Only Know Men*, by Patrick Henry, to my attention. This book tells the story of the people living in and around the French Village of Le-Chambon-sur-Lignon during the Holocaust². The book recounts the fact that the inhabitants of this village are credited with saving the lives of more than 3,500 Jewish refugees, but the truly remarkable aspect of this is that the villagers performed their rescue mission through entirely non-violent means. The book makes it clear that the people who were engaged in this rescue were ordinary people, committed to non-violence and equally committed to aiding those in need. As Henry notes, “[w]hereas others have spoken perceptively of the ‘banality of evil’ during the Holocaust, we can speak of the ‘ordinariness of goodness’...” (Henry 24).

² This book in fact is the second book to be dedicated to the same subject. It is an attempt to build upon Philip Hallie’s book *Lest Innocent Blood Be Shed*.

It is hard to choose an aspect of the actions of these villagers and single it out as particularly deserving of admiration. They performed their rescue mission in a situation in which the “Vichy government created a climate of extensive collaboration and accommodation with the Germans and where anti-Semitism and informing on Jews was deemed not only acceptable but patriotic” (Henry 12), so their very act of standing against such rampant anti-Semitism is deserving of awe. Add to this the fact that their rescue attempt came at great personal risk to the villagers. But I cannot help but feel the weight of the fact that this resistance came without sacrifice to their ideals of non-violence³. I find this example to be a helpful reminder when people charge Kant with being indifferent to tragic circumstances or take the view that his ethics is implausible in the face of evil.

It is in fact with some irony that the example of the murderer at the door is often modernized in its retelling so as to portray the image of a Nazi soldier coming to the home of someone hiding a Jewish refugee. The implication of such a version of the example is that moral standards must simply give way in the face of evil, and nothing epitomizes evil more than the actions of the Nazi regime in the Holocaust. But the example of the villagers of Le Chambon-sur-Lignon offers a different lesson: we can fight evil without violating our moral ideals. It is unfortunately the case that the tension between preventing a tragic outcome and adhering to the moral law often portrays the situation as having only two options: adhering to the moral law at the risk of allowing tragedy to occur, or using ‘common-sense’ and rejecting the choice of a strict adherence

³ Henry makes the point that, while this ethic of non-violence was for many villagers rooted in their Protestant faith, not all of the villagers were Protestant or even religious.

to moral principles for their own sake. But examples such as this village in France show us that this choice often is a false-dichotomy, and adherence to the moral law can in fact be compatible with the creation of a more moral world.

This is not to say that it is always possible to respond to immoral agents or to act so as to prevent a tragic outcome while still remaining true to the moral law. In our world it is the case that sometimes this dichotomy is not a false one at all but instead a very real one. And it is impossible, I believe, to suggest that Kant can resolve these conflicts altogether. We will sometimes be in the position of having to choose between an adherence to the dictates of the moral law and the desire to prevent a tragedy, and the only comfort Kant can give us is to argue that an abandonment of the moral law in order to prevent such tragedies is not worth the price (despite the obvious instinct that the contrary is true). The power of Kant's approach to morality is that it makes a connection between two elements of moral philosophy that are often portrayed as existing in conflict with one another. On the one hand it provides us with a clear guide to our moral conduct, which commands categorically. On the other, it provides us with a vision of a moral world we can work toward and argues that we are entitled to believe that such a world is possible. Kant connects these two because he argues that the one is the path to the other, and that the two notions are in fact inseparable. I have argued in this dissertation that commentators have often failed to recognize one or the other of these components to Kant's ethical thought. Either Kant is portrayed as entirely formalistic, or we are presented with a view of Kant that privileges the status of the Kingdom of Ends or the Highest Good in such a way that we lose sight of the connection between those concepts

and the moral law. Both of these mischaracterizations lead to views about how Kant 'should have handled' cases of "evil circumstances," and in so doing fail to give proper weight to the way he *did* handle them.

BIBLIOGRAPHY

- Anderson-Gold, Sharon. "Kant's Ethical Commonwealth: The Highest Good as Social Goal." International Philosophical Quarterly. 26.1 (1986): 23-32.
- . Unnecessary Evil: History and Moral Progress in the Philosophy of Immanuel Kant. Albany: State University of New York Press, 2001.
- Atwell, John E. Ends and Principles in Kant's Moral Thought. Boston: Martin Nijhoff Publishers, 1986.
- Auxter, Thomas. "The Unimportance of Kant's Highest Good." Journal of the History of Philosophy. 17 (1979): 121-134.
- Barnes, Gerald. "In Defense of Kant's Doctrine of the Highest Good." Philosophical Forum. 2. (Summer, 1971): 446 – 458.
- Beck, Lewis White. A Commentary on Kant's Critique of Practical Reason. Chicago: University of Chicago Press, 1960.
- Bielefeldt, Heiner. Symbolic Representation in Kant's Practical Philosophy. Cambridge, UK: Cambridge University Press: 2003.
- Billy Budd. Dir. Peter Ustinov. Perf. Robert Ryan, Peter Ustinov, Terence Stamp. DVD. Warner Brothers, 1962.
- Doore, Gary. "Contradiction in the Will." Kant-Studien: Philosophische Zeitschrift der Kant-Gesellschaft 76.2 (1985): 138-151.
- Dusing, Klaus, "Das Problem Des Hochsten Gutes in Kants Praktischer Philosophie." Kant Studien: Philosophische Zeitschrift der Kant-Gesellschaft. 62 (1971): 5 - 29.
- Engstrom, Stephen. "The Concept of the Highest Good in Kant's Moral Theory." Philosophy and Phenomenological Research. 52.4 (December, 1992): 747 – 780.
- Frankena, William. Ethics. 2nd ed. Engle Wood Cliffs, N.J.: Prentice Hall, 1973.

- Freeman, et. al. "Lying for Patients: Physician Deception of Third Party Payers." Archives of Internal Medicine. 159.19 (October, 1999): 2263-2270.
- Galvin, Richard Francis. "Ethical Formalism: The Contradiction in Conception Test." History of Philosophy Quarterly 8.4 (1991): 387-408.
- Gregor, Mary. Laws of Freedom: A Study of Kant's Method of Applying the Categorical Imperative in the Metaphysik de Sitten. New York: Barnes and Noble, 1963.
- Guyer, Paul. "Ends of Reason and Ends of Nature: The Place of Teleology in Kant's Ethics." Journal of Value Inquiry. 36 (2002): 161 – 186.
- . Kant on Freedom, Law, and Happiness. Cambridge: Cambridge University Press, 2000.
- . "The Possibility of the Categorical Imperative." Philosophical Review. 104.3 (July, 1995): 353 – 385.
- . "The Value of Reason and the Value of Freedom." Ethics. 109.1 (Oct., 1998): 22 – 35.
- Harris, Nigel G.E. "Kantian Duties and Immoral Agents." Kant-Studien: Philosophische Zeitschrift der Kant-Gesellschaft. 83.3 (1992): 336 - 343.
- Harrison, Jonathan. "Categorical Imperative." Philosophical Quarterly 8.33 (1958): 360-364.
- . "Kant's Examples of the First Formulation of the Categorical Imperative." Philosophical Quarterly 7.26 (1957): 50-62.
- Held, Virginia. Ethics of Care: Personal, Political, and Global. New York: Oxford University Press, 2006.
- Henry, Patrick. We Only Know Men: The Rescue of Jews in France during the Holocaust. Washington, D.C.: Catholic University of America Press, 2007.
- Herman, Barbara. Morality as Rationality: A Study of Kant's Ethics. New York: Garland Publishing, 1990
- . "Murder and Mayhem: Violence and Kantian Casuistry." Monist: An International Quarterly Journal of General Philosophical Inquiry 72.3 (1989): 411-431.

- . "Mutual Aid and Respect for Persons." Ethics 94.4 (1984): 577-602.
- . Practice of Moral Judgment. Cambridge, MA: Harvard University Press, 1993.
- Kant, Immanuel. "Critique of Practical Reason." Trans. Mary J. Gregor. Immanuel Kant: Practical Philosophy. Ed. Mary J. Gregor. Cambridge: Cambridge University Press, 1996.
- . Critique of Pure Reason. Trans. Norman Kemp Smith. Boston: MacMillan, 1965.
- . Critique of the Power of Judgment. Trans. Paul Guyer and Eric Matthews. Ed. Paul Guyer. Cambridge: Cambridge University Press, 2000.
- . "The End of All Things." Perpetual Peace and Other Essays. Trans. Ted Humphrey. Indianapolis: Hackett Publishing Company, 1983.
- . "Groundwork of The Metaphysics of Morals." Trans. Mary J. Gregor. Immanuel Kant: Practical Philosophy. Ed. Mary J. Gregor. Cambridge: Cambridge University Press, 1996.
- . "Idea for a Universal History With a Cosmopolitan Intent." Perpetual Peace and Other Essays. Trans. Ted Humphrey. Indianapolis: Hackett Publishing Company, 1983.
- . Lectures on Ethics. Trans. Peter Heath. Ed. Peter heath and J.B. Schneewind. Cambridge: Cambridge University Press, 1997.
- . "Lectures on the Philosophical Doctrine of Religion." Trans. Allen W. Wood. Religion and Rational Theology. Ed. Allen W. Wood and George di Giovanni. Cambridge: Cambridge University Press, 1996.
- . "The Metaphysics of Morals." Trans. Mary J. Gregor. Immanuel Kant: Practical Philosophy. Ed. Mary J. Gregor. Cambridge: Cambridge University Press, 1996.
- . "On the Proverb: That Might Be True in Theory But It Is Of No Practical Use." Perpetual Peace and Other Essays. Trans. Ted Humphrey. Indianapolis: Hackett Publishing Company, 1983.
- . "To Perpetual Peace: A Philosophical Sketch." Perpetual Peace and Other Essays. Trans. Ted Humphrey. Indianapolis: Hackett Publishing Company, 1983.
- . "Religion Within the Boundaries of Mere Reason." Trans. George di Giovanni. Religion and Rational Theology. Ed. Allen W. Wood and George di Giovanni. Cambridge: Cambridge University Press, 1996.

- Kleingeld, Pauline. "What Do the Virtuous Hope For? Re-reading Kant's Doctrine of the Highest Good." Proceedings of the Eight International Kant Congress. 1 (1995): 91-112.
- Korsgaard, Christine M. Creating the Kingdom of Ends. New York: Cambridge University Press, 1996.
- Kraft, Michael. "Kant's Theory of Teleology." Immanuel Kant: Critical Assessments Vol. IV. Ed. Ruth F. Chadwick and Clive Cazeaux. London and New York: Routledge, 1992.
- Lindstedt, David. "Kant: Progress in Universal History as a Postulate of Practical Reason." Kant-Studien: Philosophische Zeitschrift der Kant-Gesellschaft. 90 (1999): 129 – 147.
- Louden, Robert B. Kant's Impure Ethics. Oxford: Oxford University Press, 2000.
- Luban, David. "Liberalism, Torture, and the Ticking Bomb." Virginia Law Review. 91.6 (2005): 1425-1461
- McNair, Ted. "Universal Necessity and Contradictions in Conception." Kant-Studien: Philosophische Zeitschrift der Kant-Gesellschaft. 91 (2000) 25 – 43.
- Melnick, Arthur. "Kant's Formulations of the Categorical Imperative." Kant Studien: Philosophische Zeitschrift der Kant-Gesellschaft. 93 (2002): 291 – 308.
- Melville, Herman. Billy Budd, Sailor. Ed, Harrison Hayford and Merton M. Sealts Jr. Chicago: University of Chicago Press, 1962.
- Millgram, Elijah. "Does the Categorical Imperative Give Rise to a Contradiction in the Will?" Philosophical Review 112.4 (2003): 525-560.
- Moore, Jennifer. "Kant's Ethical Community." Journal of Value Inquiry. 26 (1992): 51 – 72.
- Nell, Onora. Acting on Principle: An Essay on Kantian Ethics. New York: Columbia University Press, 1975.
- O'Neill, Onora. "Universal Laws and Ends in Themselves." Monist: An International Journal of General Philosophical Inquiry 72.3 (1989): 341-361.

- Paton, H.J. Categorical Imperative: A Study in Kant's Moral Philosophy. 5th ed. New York: Hutchinson, 1965.
- Phillips, Michael. "Reflections on the Transition from Ideal to Non-Ideal Theory." Nous. 19.4 (Dec., 1985): 551 – 570.
- Plato. Crito. in Five Dialogues: Euthyphro, Apology, Crito, Meno, Phaedo. 2nd Ed. Trans. G.M.A. Grube and John M. Cooper. Indianapolis: Hackett Publishing, 2002.
- Potter, Nelson. "Kant on Ends That Are at the Same Time Duties". Pacific Philosophical Quarterly. 66 (1985): 78-92.
- . "Maxims in Kant's Moral Philosophy." Philosophia: Philosophical Quarterly of Israel 23 (1994): 59-90.
- . "Paton on the Application of the Categorical Imperative." Kant-Studien: Philosophische Zeitschrift der Kant-Gesellschaft 64.3 (1973): 411-422.
- . "What is Wrong With Kant's Four Examples." Journal of Philosophical Research 18 (1993): 213-229.
- Rawls, John. A Theory of Justice. Cambridge, MA: Belknap Press of Harvard University Press, 1971.
- Reath, Andrews. "Categorical Imperative and Kant's Conception of Practical Rationality." Monist: An International Journal of General Philosophical Inquiry. 72.3 (1989): 384-410.
- . "Two Conceptions of the Highest Good in Kant." Journal of the History of Philosophy. 26 (October, 1988): 593 – 619.
- Silber, John R. "The Importance of the Highest Good in Kant's Ethics." Ethics. 73.3 (April, 1963): 179-197.
- . "Kant's Conception of the Highest Good as Immanent and Transcendent." The Philosophical Review. 68.4 (October, 1959): 469-492.
- Simmons, Lance. "Kant's Highest Good: Albatross, Keystone, Achilles Heel." History of Philosophy Quarterly. 10.4 (October, 1993): 355 – 368.
- Singer, Marcus. "The Categorical Imperative." The Philosophical Review. 63. 4 (Oct., 1954): 577-591.

- . Generalization in Ethics: an Essay in the Logic of Ethics, with the Rudiments of a System of Moral Philosophy. New York: Knopf, 1961.
- Stratton-Lake, Phillip. "Formulating Categorical Imperatives." Kant-Studien: Philosophische Zeitschrift der Kant-Gesellschaft. 83 (1993): 317 – 340.
- Ward, Keith. Development of Kant's Ethics. New York: Humanities Press, 1972.
- . "Kant's Teleological Ethics." Immanuel Kant: Critical Assessments Vol. III. Ed. Ruth F. Chadwick. London and New York: Routledge, 1992.
- Werner, et. al. "Lying to Insurance Companies: The Desire to Deceive among Physicians and the Public." American Journal of Bioethics. 4.4 (2004): 1-7.
- Wike, Victoria. Kant on Happiness in Ethics. Albany: SUNY Press, 1995.
- Williams, T.C. The Concept of the Categorical Imperative: A Study of the Place of the Categorical Imperative in Kant's Ethical Theory. Oxford: Clarendon Press, 1968.
- Wood, Allen W. Kant's Ethical Thought. Cambridge: Cambridge University Press, 1999.
- . Kant's Moral Religion. Ithaca: Cornell University Press, 1970.

VITA

Before attending Loyola University of Chicago, John Garcia attended Boston College, where he earned a Bachelor of Arts in Philosophy in 1997. In 1997-1998 he worked at Ateneo de Manila University, in the Philippines, as an instructor of Philosophy and as an instructor of History. From 1999-2001 he attended Northern Illinois University, where he earned a Master of Arts in Philosophy.

While at Loyola, John worked as a graduate assistant at the Loyola University Chicago Center for Ethics. Additionally, he worked as an adjunct professor at Kishwaukee Community College, Harper College, Oakton Community College, the School of the Art Institute of Chicago, and Loyola University Chicago.

John is currently an Assistant Professor of Philosophy at Harper College, in Palatine Illinois. He lives in Oak Park, Illinois, with his wife Larissa.